Cleavage of DNA by Polyamide-*seco*-CBI Conjugates

Thesis by
Aileen Yulin Chang

In Partial Fulfillment of the Requirements
for the Degree of
Doctor of Philosophy

California Institute of Technology
Pasadena, California
2001
(Submitted February 23, 2001)

To Jeremy and My Parents

Acknowledgements

I am grateful to my family for all their encouragement during my stay at Caltech. My husband Jeremy has helped me get through many bad days, and to help me see that I'll still survive, even if none of my reactions work for the rest of my life. I'll always be thankful for his patience and wisdom when I wasn't so keen on receiving good advice. I hope that as we grow together, that I'll be maturing in those areas that he's so strong in: discipline, wisdom, and having a good perspective on life. My parents have made so many sacrifices for me over the courses of their lives. They neglected to give themselves so many material things as they raised my brother, sister, and I. It's gratifying for me to see that in their retirement, they're able to enjoy life through travel and relaxation. Certainly they deserve it!

There are so many friends that have meant so much to me in my stay at Caltech. My former roommate Deanna Zubris, and our 'Chick Lunch' gang, Elizabeth Krider and Claudine Chen have provided sympathetic ears and encouragement over dessert at the Red Door or other places. 'Grad Group' of Caltech Christian Fellowship has been an unending source of encouragement and satisfying Bible study discussions. I've appreciated the many things they have lifted up in prayer for myself and my family. I'll miss the many friendships that Jeremy and I have developed over the years from this group, and I believe that we'll see a great generation of Christian scientists, engineers, professors, and their spouses that count grad group in their roots.

There are many folks from the Dervan group to thank. My labmates Clay Wang, Ryan Bremer, and Nick Wurtz have been encouraging through lots of miserably failed reactions. I hope that someday I can learn to be as gracious a host as Clay and Tammy are. Dervanites everywhere, myself included, have been impressed by their hospitality and generosity. With his determination, I know that Clay will be successful at whatever he tries to do. It's been a privilege to meet so many smart people from Grinnell College: Ryan and Lis, and Nick and Heather. Ryan and Nick have been exemplary labmates, always having insightful comments, as well as kind words when necessary. I'll miss seeing Alexander Wurtz, but I know his parents will do a great job to make sure he turns into a wonderful kid.

I'll always be thankful to older (and wiser) group members who helped me along the way. Eldon Baird, Sarah White, Jason Sczewczyk (chef-chik) were especially helpful when I first arrived, and I've always appreciated their patience and helpfulness. Jim Turner is the smartest undergrad/post doc that I've ever met. David Herman (alternate spellings include Hairmone, Hermaan, Hermán) has always provided lots of laughs, even without caffeine. Tom Minehan offered lots of good advice about chemistry and life, and could always put a smile on your face with his 'iferous-isms.' Ulf Ellervik is the craziest Swede I know, and his zest for life and chemistry are unmatched around the globe. I've enjoyed getting to know Doan Nguyen and Jason Belitsky, who with their insight and intelligence mixed in with a good sense of humor have provided intriguing conversation. Victor Rucker is a special character, I'll miss our conversations and seeing 'the lab dance.'

Without Peter Dervan, there would be Dervan group, and I'll always be thankful that he took me in as a first year. In many ways, I feel unworthy of being a Dervan group PhD. He has set a high standard for research, and I'm not sure I've met those. His constant enthusiasm and passion for our work has always been an example for me to follow. I'm also grateful to my committee, Professors Bill Goddard, Doug Rees, David Tirrell, and Erick Carreira. My getting here was not always so smooth, but I appreciate the efforts that they invested in me along the way.

Many Caltech staff members have made my stay here much easier and more pleasant: Dian Buchness, Margot Hoyt, Lynne Martinez, Tom Dunn, Mass Spec Facility (Gary Hathaway, Jie Zhou, Felicia Rusnak), Cora Carriedo at VWR and the Student Health Center. I'm sure that I'm missing folks to thank, and for that I must ask your forgiveness. Every person I've had the chance to meet at Caltech has touched me in a special way. Thanks!

Abstract

Small molecules that bind to any predetermined DNA sequence in the human genome are potentially useful tools for molecular biology and human medicine. Polyamides containing N-methylimidazole (Im) N-methylpyrrole (Py) are cell permeable small molecules that bind DNA according to a set of "pairing rules" with affinities and specificities similar to many naturally occurring DNA binding proteins. Py-Im polyamides offer a general approach to the chemical regulation of gene expression. We demonstrate here that polyamide containing a DNA alkylating moiety *seco*-CBI can specifically direct sequence specific DNA alkylation. We can also control the strand of DNA that is alkylated, depending on the enantiomer of *seco*-CBI used and the orientation of the polyamide relative to the alkylation site (Chapter 2). This class of molecules has been applied to a gene repair system in collaboration with the Baltimore group at Caltech (Chapter 3). Also reported are additional *seco*-CBI polyamide conjugates synthesized to study other systems (HIV-1 and COX-2) (Appendix 1).

Table of Contents

List of Figures

Chapter 1 page

Chapter 2 page

Chapter 3 page

Appendix 1 page

CHAPTER ONE

Introduction

Background

Expression of Genetic Information from DNA

DNA is the universal storage material for genetic information in all living organisms. The flow of genetic information starts with DNA, which is transcribed into RNA, and the RNA is then translated to proteins. DNA serves as the permanent storage material in the cell, while RNA is a transient copy of this genetic information and is synthesized as necessary. Proteins serve to perform specific functions such as signaling or maintenance in the cell, and like RNA, are often synthesized as needed and then degraded. It is remarkable, then, that all the genetic information essential to sustaining living organisms is present in the form of DNA and that copies of this DNA are present in every living cell.

Specific protein-DNA interactions are fundamental pieces of cell differentiation. For although almost all cells contain identical genetic information, the selective transcription of certain genes but not others allows a neural cell in the brain to function differently than an epithelial cell in the skin.[1] Transcription factors are DNA binding proteins that regulate transcription by recruiting the necessary transcriptional machinery. Typically, the promoter region of a gene contains the recognition sites of the required transcription factors. Upon binding, the transcription factor recruits a network of more than fifty proteins that are required to initiate active transcription. This multi-protein complex has a molecular weight of more than two million, and contains all the proteins necessary for activated transcription to occur.[2] The process of transcription elongation is just as complex, as more than 20 protein subunits that comprise RNA polymerase II must track along the DNA and synthesize the complementary RNA. Unlike in DNA synthesis,

transcription elongation occurs in fits and starts, with elongation factors necessary to start the polymerase complex again once it has paused.[3,4]

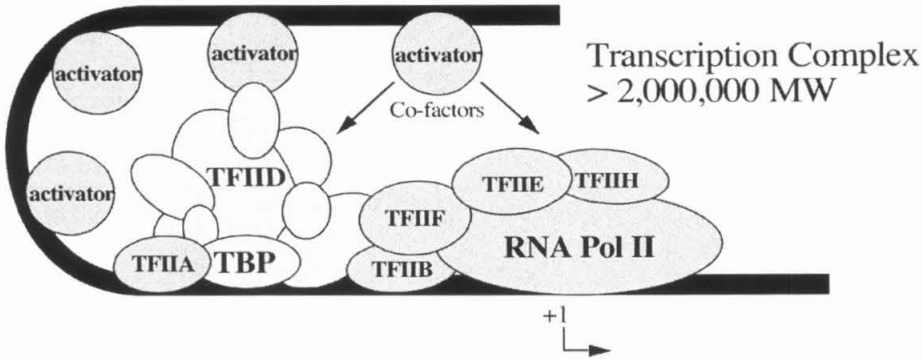


Figure 1.1. Model of protein regulation of gene transcription.

DNA as a paradigm for small molecule recognition

Because DNA plays a fundamental role in gene expression in the cell, a general code for sequence specific DNA recognition is desirable for the design of new drugs that manipulate gene expression.[5] Much of the work in this field is based on protein:DNA interactions which are characterized by hydrogen bonds, van der Waals interactions and electrostatic interactions. To date, no all-purpose code of amino acid/DNA recognition has been developed due to the diverse recognition and structural motifs found among DNA-binding proteins.[6]

To develop general rules for sequence specific DNA recognition by small molecules, a careful examination of DNA structure is required. Double helical DNA consists of two antiparallel strands that fit together in a right-handed helix.[7] DNA can be characterized as having three domains: the sugar/phosphate backbone, the major groove, and the minor groove. The deoxyribose and phosphate esters form a unique charge pattern along the helix backbone. Protein side chains often make electrostatic

interactions with the sugar/phosphate backbone, but such interactions tend not to be sequence specific.

The nature of DNA helix dictate the characteristic structures of the major and minor grooves, the major groove being wider and shallower, and the minor groove being narrower and deeper. The edges of the base pairs in the Watson-Crick hydrogen bond pairings result in the unique hydrogen bonding faces of the major and minor groove.[7] For a T•A base pair, in the major groove there is a pattern of hydrogen bond acceptor (lone pairs on O4 of thymine), and hydrogen bond donor (NH of exocyclic amine of adenine). The minor groove of a T•A base pair forms a pattern of two hydrogen bond acceptors (O2 of thymine and N3 of adenine). Likewise, for a C•G base pair, the major groove can be characterized as having a hydrogen bond donor (NH of exocyclic amine of cytosine) and a hydrogen bond acceptor (O6 of guanine). The minor groove of a C•G base pair has a hydrogen bond acceptor (O2 of cytosine), hydrogen bond donor (NH of exocyclic amine of guanine) and a hydrogen bond acceptor (N3 of guanine) (Figure 1.3). In the minor groove, it is the exocyclic amine of guanine that primarily distinguishes an A•T base pair from a G•C base pair.

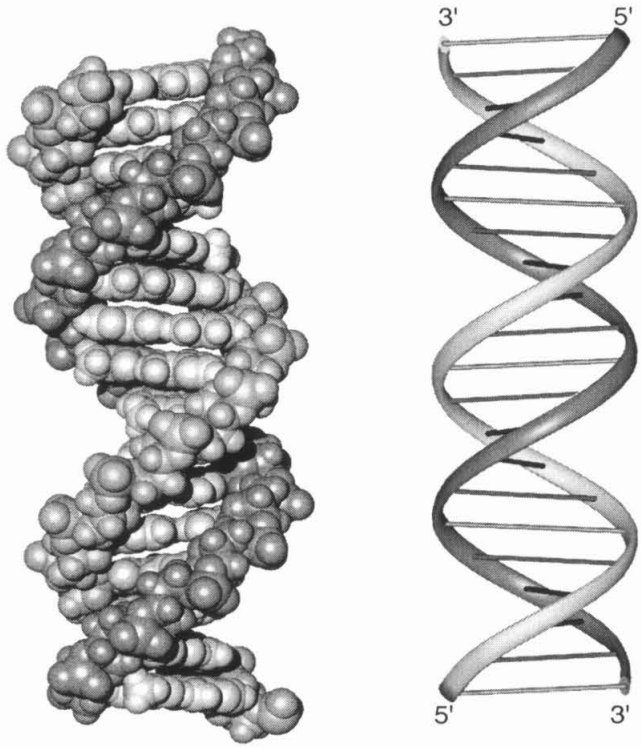


Figure 1.2. B-form double helical DNA. Antiparallel strands are indicated in dark and light gray. (left) space filling CPK model. (right) ribbon representation.

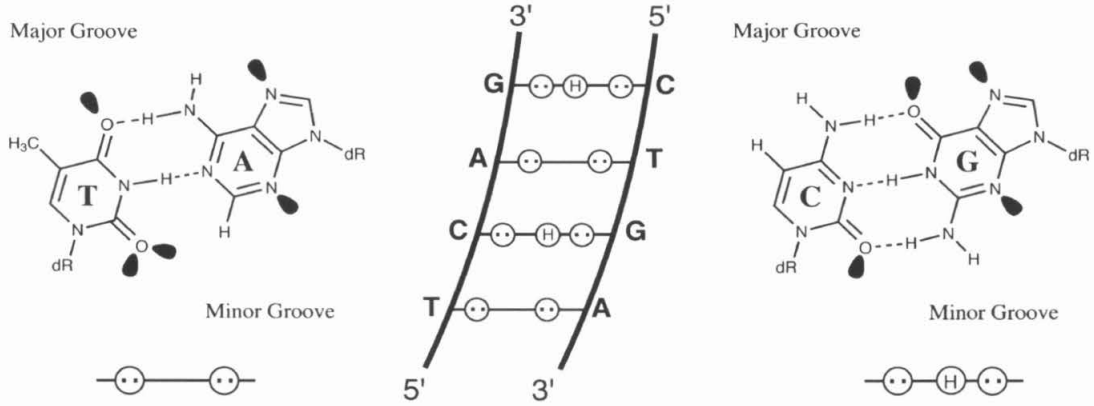


Figure 1.3. A schematic model for recognition of the minor groove, with hydrogen bond donors represented as (H) and hydrogen bond acceptors represented as two dots.

Sequence Specific DNA binding by Distamycin and 2-Imidazole Netropsin

In addition to proteins, there are many classes of natural products that are able to bind to DNA in a sequence specific manner. Some examples of such molecules are chromomycin, distamycin, actinomycin D, and calicheamicin oligosaccharide (Figure 1.4).[8-11] These molecules either bind in the minor groove, or intercalate between the base pairs. Because of their size they have short DNA recognition sequences, and their interactions could be compared to those of protein side chains.

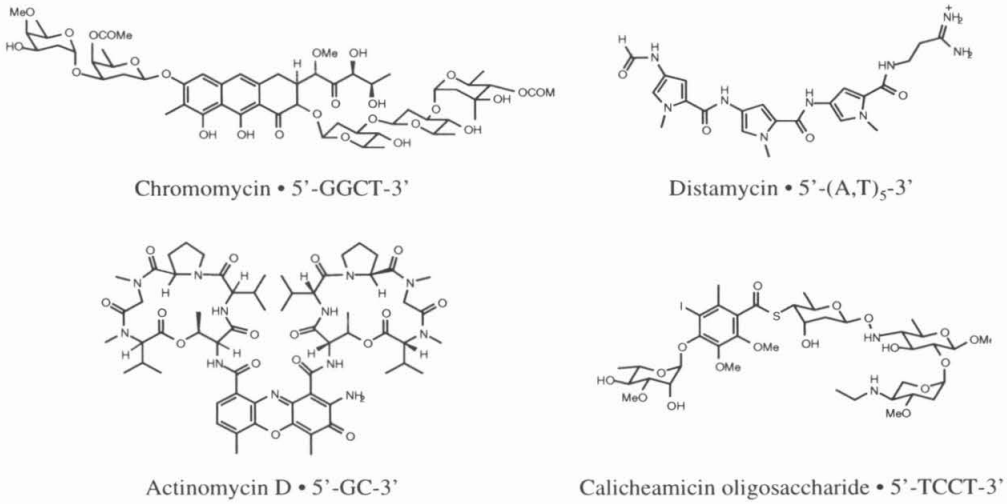
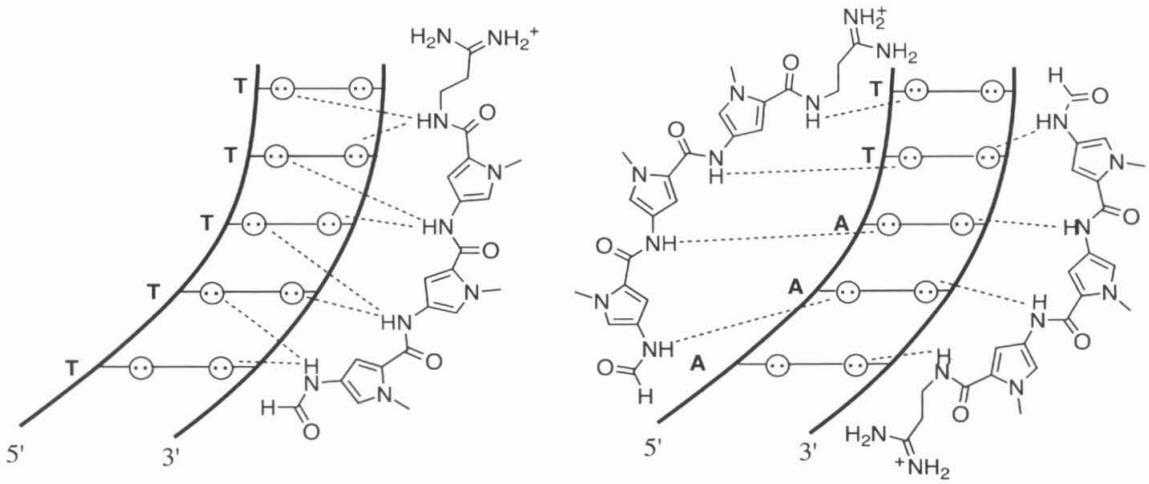


Figure 1.4. Chemical structures of natural products that bind DNA.

Distamycin is distinct from most DNA binding natural products only by its simplicity. It consists of three N-methyl carboxamide units and contains no chiral centers. This crescent shaped molecule binds in the minor groove of DNA, preferentially at A-T rich regions. The crystal structure of distamycin bound to a DNA duplex showed that the carboxamides make specific hydrogen bonds to the bases in the minor groove.[11] The Dervan group has pursued distamycin as the starting point to base the discovery of small molecules that regulate transcription.[12]

An early lead in attempts to generate a general code for DNA recognition was the discovery of 2-imidazole netropsin. It had been suggested that the imidazole ring of this compound would make a specific hydrogen bond to a G•C base pair and that 2-imidazole netropsin would target the sequence 5'-GWWW-3'. Footprinting experiments showed that unexpectedly, 2-imidazole netropsin instead bound the site 5'-WGWCW-3'. The two-fold symmetry of the binding site was puzzling since it suggested that there was not a single binding orientation for 2-imidazole netropsin. Earlier studies of distamycin/DNA complexes gave an interesting solution to this discrepancy. Distamycin was found to have two very distinct binding modes. In the first, a single molecule of distamycin binds in the center of the minor groove of A-T rich DNA. The amide hydrogens form bifurcated hydrogen bonds to the N3 of adenine and O2 of thymine on the floor of the minor groove. In the second binding mode, distamycin binds as an antiparallel, side-by-side dimer in the minor groove. Instead of the bifurcated hydrogen bonds of the 1:1 complex, each strand of DNA forms hydrogen bonds with one molecule of distamycin. Using this 2:1 binding model for 2-imidazole netropsin explained why its preferred binding site is symmetrical (because it binds as a side-by-side antiparallel dimer) and does not match the predicted site.[13,14]

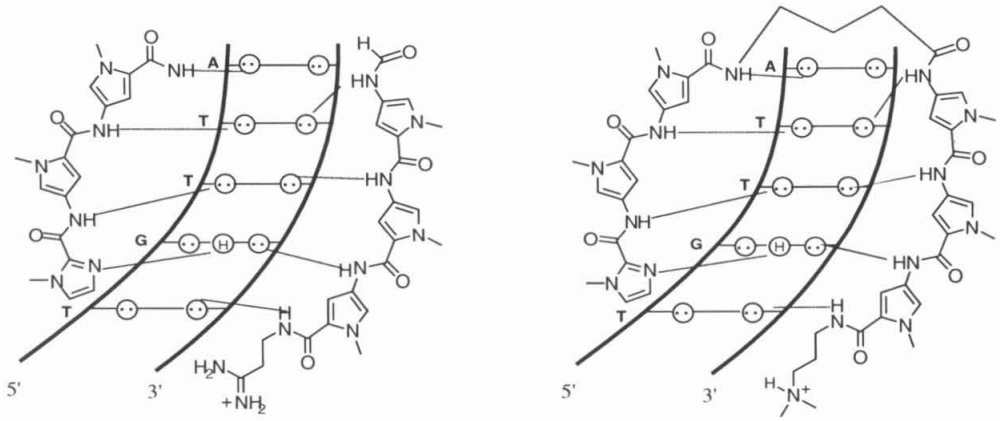


1:1 Distamycin•DNA Complex *2:1 Distamycin•DNA Complex*

Figure 1.5. A schematic representation of recognition of A,T rich sequences in the minor groove by 1:1 and 2:1 complexes of Distamycin.

The Hairpin Motif of Pyrrole-Imidazole Polyamides

The affinity of the dimer 2-imidazole netropsin was still modest (micromolar binding constants) in comparison to DNA binding proteins (nanomolar or higher binding constants). To decrease the entropic factors influencing binding, two molecules of 2-imidazole netropsin were connected in a head to tail fashion with an alkyl linker, γ-amino-butyric acid (γ).[15] The resulting 'hairpin'-like structure yielded a compound with increased affinity for the binding site 5'-WGWCW-3' of ~100 fold. Connecting the two distamycin-like components also allowed for the design of heterodimeric polyamides to target asymmetric sequences. As an example, the compound ImPyPy-γ-PyPyPy was found to target the sequence 5'-WGWWW-3'. The hairpin motif for this compound was found to increase the binding affinity by 400-fold (Figure 1.6).



ImPyPy-Dp / PyPyPy • TGTTA
$K_a \sim 2 \times 10^5\ M^{-1}$

ImPyPy-γ-PyPyPy-Dp • TGTTA
$K_a = 8 \times 10^7\ M^{-1}$

Figure 1.6. Example of the hairpin polyamide motif.

Pairing Rules for Pyrrole-Imidazole Polyamides

The results of 2-imidazole netropsin binding to the site 5'WGWCW-3' as an antiparallel dimer and the development of the hairpin motif allowed us to design a set of pairing rules for recognition by pyrrole (Py) and imidazole (Im) amino acids[16] (Figure 1.7). A Py/Im pair targets a C•G base pair while an Im/Py pair targets a G•C base pair.[17,18] The basis for discrimination of a G•C base pair is from the formation of a hydrogen bond between imidazole N3 and the exocyclic amine of guanine.[19,20] A Py/Py pair is partially degenerate and recognizes both A•T and T•A base pairs.[13,17,18,21,22] High resolution NMR and crystal structures have confirmed that specific hydrogen bonds are made from the carboxamide nitrogens to the base pairs and the imidazole nitrogen to guanine.

To address the question of A•T/T•A degeneracy by a Py/Py pair, N-methyl-3-hydroxypyrrole (Hp) was synthesized and incorporated into a polyamide.[23,24] Hp places a

hydroxy group in the asymmetric cleft between a thymine and adenine base pair. Thus, an Hp/Py pair recognizes the sequence T•A and a Py/Hp pair recognizes an A•T base pair (Figure 1.7). The binding of Hp has also been confirmed by a high resolution crystal structure, and shows that the hydroxy group of Hp binds in the asymmetric cleft as designed and also makes two hydrogen bonds to O2 of thymine.[25,26]

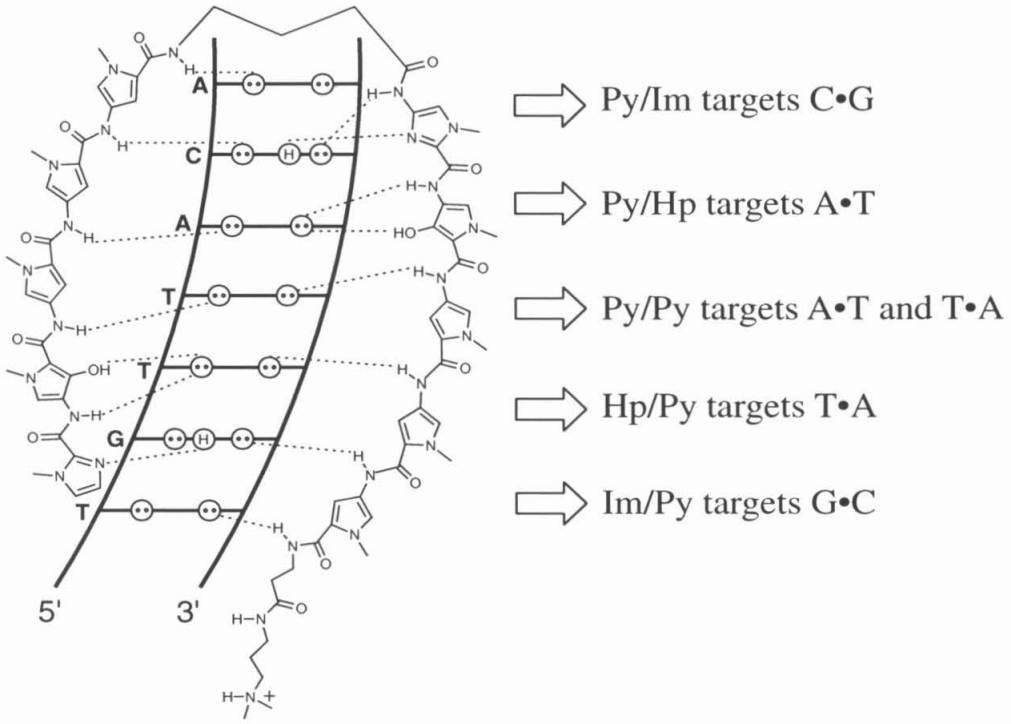


Figure 1.7. Binding model for the complex formed between ImHpPyPyPy-γ-ImHpPyPyPy-β-Dp and the DNA duplex 5'-TGTTACA-3'. Hydrogen bonds are shown as dashed lines.

Other polyamide motifs

The development of a solid phase methodology to synthesize Py-Im polyamides has greatly facilitated the examination of other motifs and analogues of polyamides.[27] Another modification to the hairpin structures is the use of a chiral turn (Diamino-butyric acid, DABA) which influences the binding orientation of the polyamide and increases affinity and specificity.[28] Extension of the binding site size was achieved by the addition

of more *Py or Im* rings to the hairpin. In order to target sequences of more than six to eight base pairs, an aliphatic β-alanine residue (β) can be used.[29-32] The crystal structure of a Py-Im polyamide shows that the helical pitch of the polyamide is slightly greater than the pitch of the DNA helix.[20] Thus to target longer sequences, β is added to compensate for these sequence composition effects. Using these modifications, it is possible to design polyamides having subnanomolar binding constants, on par with those of DNA binding proteins. Examples of other polyamide motifs such as cycles[33,34], tandem hairpins[35], and cooperative dimers[36,37] are shown in Figure 1.8.

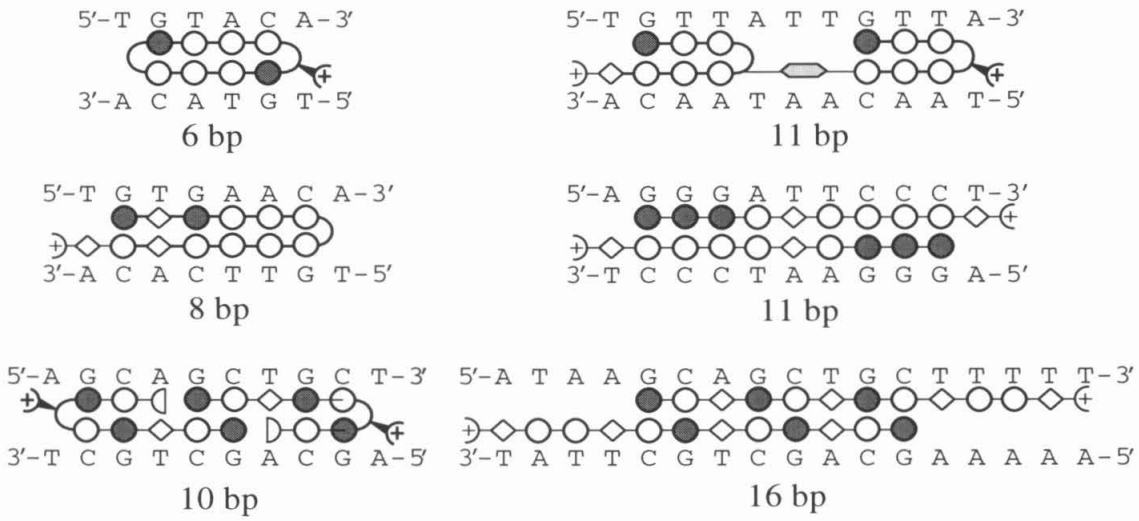


Figure 1.8. Representative motifs for polyamide:DNA recognition. Shaded and unshaded circles represent imidazole and pyrrole residues, respectively. β-alanine and the γ-turn are represented as a diamond and curved line, respectively. The plus sign represents the dimethylaminopropylamide tail or the α-amino-γ-turn.

Applications of Py-Im Polyamides in Transcription Inhibition and Activation

As a proof of principle that small molecules can regulate transcription, Py-Im polyamides were designed for use in two transcription systems. For the first system, two polyamides were designed to inhibit the binding of TATA-binding protein (TBP) and LEF-1, two essential transcription factors that bind in the promoter region of HIV-1.[29]

Individually, each was shown to inhibit HIV-1 transcription between 60% and 80%. However, together the two polyamides showed synergistic effects and reduced HIV-1 transcription to undetectable levels in cell culture. This experiment showed that Py-Im polyamides are cell permeable, and are capable of inhibiting transcription through rational design to targeted sequences.

A second system used a constructed promoter to see if polyamides were capable of activating transcription.[38] The polyamide ImPyPyPy-γ-PyPyPyPy-β-Dp was coupled to a known transcriptional activator peptide. A promoter with six match sites for the polyamide-peptide conjugate was placed 40 base pairs upstream from the TATA box. Up to 40-fold activation of transcription was observed over basal levels with these conjugates. These experiments show that polyamides can be used to manipulate a variety of transcriptional processes, and point the way to other experiments using small molecules that sequence specifically bind to DNA.

Scope of this Work

Sequence Specific DNA Alkylating Agents

Py-Im hairpin polyamides are small molecules that are cell permeable and capable of inhibiting transcription in *in vivo* systems through sequence specific DNA interactions. Thus far, the methods explored for transcription inhibition by polyamides have focused on preventing the binding of essential transcription factors in the promoter region. In unpublished work, we have found that when polyamides are targeted to the coding region of specific genes, no inhibition of transcription is observed. The noncovalent interactions between the polyamide and DNA are not strong enough to stop the progression of RNA polymerase. We were interested in addressing the question of whether bifunctional DNA

binding polyamides that covalently interact with the minor groove of DNA could inhibit polymerase elongation. The generation of a sequence specific DNA alkylating agent would introduce a new class of gene specific 'knockout' agents that would be useful for a variety of biological disciplines.

Chapter 2 describes efforts to develop a sequence specific alkylating agent using a hairpin polyamide-*seco*-CBI conjugate. CBI is an analogue of the natural product CC-1065, and alkylates DNA in the minor groove at N3 of adenine. Hairpin polyamides modified with *seco*-CBI off the γ-amino-butyric acid turn were synthesized to deliver the alkylating moiety to adenines adjacent to the polyamide binding site. Applications of these polyamide conjugates are discussed in chapter 3, through the design of polyamide conjugates to study gene repair processes. Appendix 1 describes other polyamide-*seco*-CBI conjugates that were characterized to study COX-2 and HIV.

In this work, the function of hairpin polyamides has been extended to include sequence specific covalent modification of DNA in the minor groove. The generality of the pairing rules and the knowledge at hand concerning transcription elongation and DNA repair processes make sequence specific DNA alkylation a useful approach for the study of these systems.

References

1. Roeder, R. G. The role of general initiation factors in transcription by RNA polymerase II. *Trends Biochem.Sci.* **21**, 327-335 (1996).
2. Tjian, R. Molecular Machines That Control Genes. *Sci. Am.* **272**, 54-61 (1995).
3. Conaway, J. W. & Conaway, R. C. Structural biology - Light at the end of the channel. *Science* **288**, 632-633 (2000).
4. Cramer, P. et al. Architecture of RNA polymerase II and implications for the transcription mechanism. *Science* **288**, 640-649 (2000).
5. Gottesfeld, J. M., Neely, L., Trauger, J. W., Baird, E. E. & Dervan, P. B. Regulation of gene expression by small molecules. *Nature* **387**, 202-205 (1997).
6. Pabo, C. O. & Sauer, R. T. Transcription Factors - Structural Families and Principles of DNA Recognition. *Annu. Rev. Biochem.* **61**, 1053-1095 (1992).
7. Watson, J. D. & Crick, F. H. C. *Nature* **171**, 737 (1953).
8. Paloma, L. G., Smith, J. A., Chazin, W. J. & Nicolaou, K. C. Interaction of Calicheamicin with Duplex DNA - Role of the Oligosaccharide Domain and Identification of Multiple Binding Modes. *J. Am. Chem. Soc.* **116**, 3697-3708 (1994).
9. Kamitori, S. & Takusagawa, F. Crystal-Structure of the 2/1 Complex between D(GAAGCTTC) and the Anticancer Drug Actinomycin-D. *J. Mol. Biol.* **225**, 445-456 (1992).
10. Gao, X. L., Mirau, P. & Patel, D. J. Structure Refinement of the Chromomycin Dimer-DNA Oligomer Complex in Solution. *J. Mol. Biol.* **223**, 259-279 (1992).

11. *Coll, M.*, Frederick, C. A., Wang, A. H. J. & Rich, A. A Bifurcated Hydrogen-Bonded Conformation in the D(A.T) Base- Pairs of the DNA Dodecamer D(CGCAAATTTGCG) and Its Complex with Distamycin. *Proc. Natl. Acad. Sci. USA.* **84**, 8385-8389 (1987).

12. Dervan, P. B. Design of Sequence-Specific DNA-Binding Molecules. *Science* **232**, 464-471 (1986).

13. Chen, X., Ramakrishnan, B., Rao, S. T. & Sundaralingam, M. Binding of 2 Distamycin-A Molecules in the Minor-Groove of an Alternating B-DNA Duplex. *Nat. Struct. Biol.* **1**, 169-175 (1994).

14. Pelton, J. G. & Wemmer, D. E. Structural Modeling of the Distamycin-a-D(CGCGAATTCGCG)$_2$ Complex Using 2D NMR and Molecular Mechanics. *Biochemistry* **27**, 8088-8096 (1988).

15. Mrksich, M. & Dervan, P. B. Design of a Covalent Peptide Heterodimer For Sequence-Specific Recognition in the Minor-Groove of Double-Helical DNA. *J. Am. Chem. Soc.* **116**, 3663-3664 (1994).

16. Dervan, P. B. & Burli, R. W. Sequence-specific DNA recognition by polyamides. *Current Opinion in Chemical Biology* **3**, 688-693 (1999).

17. Wade, W. S., Mrksich, M. & Dervan, P. B. Design of Peptides That Bind in the Minor Groove of DNA At 5'- (A,T)G(A,T)C(A,T)-3' Sequences By a Dimeric Side-By-Side Motif. *J. Am. Chem. Soc.* **114**, 8783-8794 (1992).

18. White, S., Baird, E. E. & Dervan, P. B. On the pairing rules for recognition in the minor groove of DNA by pyrrole-imidazole polyamides. *Chem. Biol.* **4**, 569-578 (1997).

19. *Geierstanger*, B. H., Mrksich, M., Dervan, P. B. & Wemmer, D. E. Design of a G-C-Specific DNA Minor Groove-Binding Peptide. *Science* **266**, 646-650 (1994).

20. Kielkopf, C. L., Baird, E. E., Dervan, P. D. & Rees, D. C. Structural basis for G•C recognition in the DNA minor groove. *Nat. Struct. Biol.* **5**, 104-109 (1998).

21. Pelton, J. G. & Wemmer, D. E. Structural Characterization of a 2-1 Distamycin a.D(CGCAAATTGGC) Complex By Two-Dimensional NMR. *Proc. Natl. Acad. Sci. USA* **86**, 5723-5727 (1989).

22. White, S., Baird, E. E. & Dervan, P. B. Effects of the A•T/T•A degeneracy of pyrrole-imidazole polyamide recognition in the minor groove of DNA. *Biochemistry* **35**, 12532-12537 (1996).

23. White, S., Szewczyk, J. W., Turner, J. M., Baird, E. E. & Dervan, P. B. Recognition of the four Watson-Crick base pairs in the DNA minor groove by synthetic ligands. *Nature* **391**, 468-471 (1998).

24. White, S., Turner, J. M., Szewczyk, J. W., Baird, E. E. & Dervan, P. B. Affinity and specificity of multiple hydroxypyrrole/pyrrole ring pairings for coded recognition of DNA. *J. Am. Chem. Soc.* **121**, 260-261 (1999).

25. Kielkopf, C. L. et al. Structural effects of DNA sequence on T•A recognition by hydroxypyrrole/pyrrole pairs in the minor groove. *J. Mol. Biol.* **295**, 557-567 (2000).

26. Kielkopf, C. L. et al. A structural basis for recognition of A•T and T•A base pairs in the minor groove of B-DNA. *Science* **282**, 111-115 (1998).

27. Baird, E. E. & Dervan, P. B. Solid phase synthesis of polyamides containing imidazole and pyrrole amino acids. *J. Am. Chem. Soc.* **118**, 6141-6146 (1996).

28. Herman, D. M., Baird, E. E. & Dervan, P. B. Stereochemical control of the DNA binding affinity, sequence specificity, and orientation preference of chiral hairpin polyamides in the minor groove. *J. Am. Chem. Soc.* **120**, 1382-1391 (1998).

29. Dickinson, L. A. et al. Inhibition of RNA polymerase II transcription in human cells by synthetic DNA-binding ligands. *Proc. Natl. Acad. Sci. USA* **95**, 12890-12895 (1998).

30. Swalley, S. E., Baird, E. E. & Dervan, P. B. A pyrrole-imidazole polyamide motif for recognition of eleven base pair sequences in the minor groove of DNA. *Chem.-Eur. J.* **3**, 1600-1607 (1997).

31. Trauger, J. W., Baird, E. E., Mrksich, M. & Dervan, P. B. Extension of sequence-specific recognition in the minor groove of DNA by pyrrole-imidazole polyamides to 9-13 base pairs. *J. Am. Chem. Soc.* **118**, 6160-6166 (1996).

32. Turner, J. M., Swalley, S. E., Baird, E. E. & Dervan, P. B. Aliphatic/aromatic amino acid pairings for polyamide recognition in the minor groove of DNA. *J. Am. Chem. Soc.* **120**, 6219-6226 (1998).

33. Cho, J. Y., Parks, M. E. & Dervan, P. B. Cyclic Polyamides for Recognition in the Minor-Groove of DNA. *Proc. Natl. Acad. Sci. USA* **92**, 10389-10392 (1995).

34. Herman, D. M., Turner, J. M., Baird, E. E. & Dervan, P. B. Cycle polyamide motif for recognition of the minor groove of DNA. *J. Am. Chem. Soc.* **121**, 1121-1129 (1999).

35. Herman, D. M., Baird, E. E. & Dervan, P. B. Tandem hairpin motif for recognition in the minor groove of DNA by pyrrole - Imidazole polyamides. *Chem.-Eur. J.* **5**, 975-983 (1999).

36. Trauger, J. W., Baird, E. E. & Dervan, P. B. Cooperative hairpin dimers for recognition of DNA by pyrrole- imidazole polyamides. *Angew. Chem.-Int. Edit. Engl.* **37**, 1421-1423 (1998).

37. Trauger, J. W., Baird, E. E. & Dervan, P. B. Recognition of 16 base pairs in the minor groove of DNA by a pyrrole-imidazole polyamide dimer. *J. Am. Chem. Soc.* **120**, 3534-3535 (1998).

38. Mapp, A. K., Ansari, A. Z., Ptashne, M. & Dervan, P. B. Activation of gene expression by small molecule transcription factors. *Proc. Natl. Acad. Sci. USA* **97**, 3930-3935 (2000).

CHAPTER TWO

Strand Selective Cleavage of DNA by Diastereomers of Hairpin Polyamide-seco-CBI Conjugates

The text of this chapter is taken in part from a published paper that was coauthored with Prof. Peter B. Dervan
(Chang, A.Y.; Dervan, P.B. *J. Am. Chem. Soc.* **2000**, 122, 4856-4864)

Abstract

Pyrrole-imidazole polyamides are synthetic ligands that bind predetermined DNA sequences with subnanomolar affinity. We report the synthesis and characterization of an eight-ring hairpin polyamide conjugated at the turn to both enantiomers of 1-(chloromethyl)-5-hydroxy-1,2-dihydro-3*H*-benz[e]indole (*seco*-CBI), an alkylating moiety related to CC-1065. Alkylation yields and specificity were determined on a restriction fragment containing six base pair match and mismatch sites. Alkylation was observed at a single adenine flanking the polyamide binding site, and strand selective cleavage could be achieved based on the enantiomer of *seco*-CBI chosen. At 1 nM concentrations of polyamide-*seco*-CBI conjugate, near quantitative cleavage was observed after 12 hours. These bifunctional molecules could be useful for targeting coding regions of genes and inhibition of transcription.

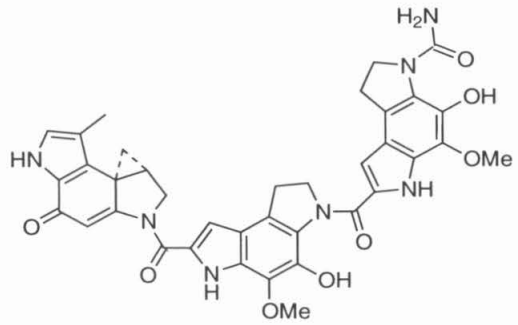
Design of Polyamide-CBI Conjugates

The design of sequence specific DNA alkylating agents requires the integration of two separate functional moieties for recognition and reaction. As described in chapter 1, hairpin Py-Im polyamides are suitable for DNA recognition due to their high affinity and specificity for the minor groove of DNA. The reactive moiety must specifically alkylate in the minor groove proximal to the hairpin polyamide target site, with covalent reaction yields that are near quantitative. In order to maximize stoichiometric reaction on the DNA, the 'electrophilic functionality' must be reactive with DNA at 37 °C, be inert in aqueous media and buffer components, and not suffer unimolecular decomposition in competition with the desired reaction with DNA. The reactive moiety of the natural product (+)-CC-1065 and its analogues meets these criteria for our design of bifunctional molecules for sequence specific alkylation of DNA.

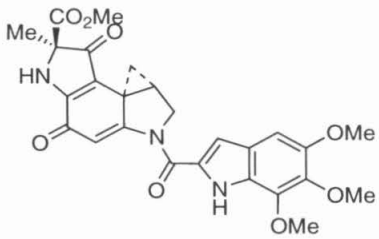
CC-1065 and Duocarmycins

CC-1065 is part of the family of duocarmycins, natural products derived from cultures of *Streptomyces zelensis*. It consists of three substituted benzopyrrole moieties linked by amide bonds. It was discovered to be a potent antitumor antibiotic by Upjohn. The mode of action was found to be the reactive cyclopropyl moiety on one of the subunits that affords sequence specific DNA alkylation at N3 of adenine.[1] This reactive moiety is typical of all members of the duocarmycin family (Figure 2.1). Alkylation at this position results in a labile base-sugar bond generating an abasic site that leads to DNA cleavage. The duocarmycins target A/T rich sequences of DNA. Unlike the case of distamycin, the duocarmycins have no amide protons with which to form hydrogen bonds to the DNA bases. Therefore, its sequence selectivity is based primarily on van

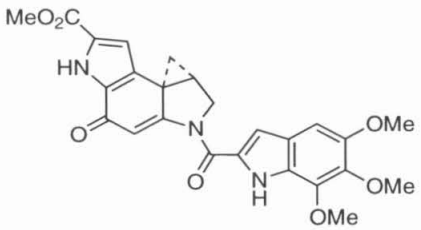
der Waals interactions that are maximized in the narrower A/T minor groove. CC-1065 once showed great promise as an antitumor agent for treatment of cancer, but its early clinical trials revealed delayed toxicity.[2,3] However, its mechanisms of action continue to be studied and pursued for the discovery of related anticancer drugs.



(+)- CC-1065



(+)- Duocarmycin A



(+)- Duocarmycin SA

Figure 2.1. CC-1065 and the family of Duocarmycins.

A debate has lingered for some time concerning the origin of catalysis for the alkylation reaction of the duocarmycins at N3 of adenine.[4] The Alkylation Site model requires that the selectivities of simple and extended analogues of CC-1065 are derived only from the cyclopropyl moiety. Noncovalent interactions are not considered to have significant effect on the alkylation selectivity. This model requires acid dependence to promote protonation of the C4 carbonyl for alkylation of N3 of adenine to take place (Figure 2.2a).

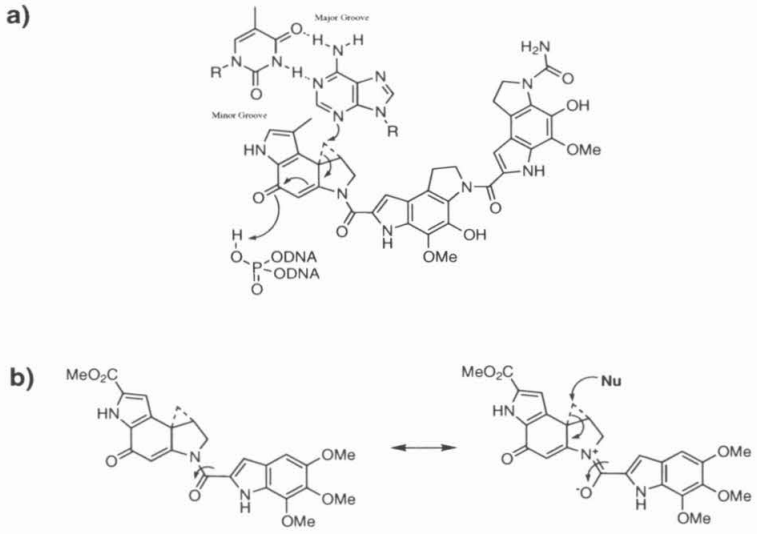


Figure 2.2. (a) Alkylation Site model. Alkylation reaction is driven by protonation of the carbonyl, presumably from a protonated phosphate of the DNA backbone. (b) Illustration of vinylagous amide stabilization in duocarmycins. Stability to nucleophiles in solution is enhanced by the conjugation of the rings systems.

Dale Boger has argued against this model and suggested the Noncovalent Binding model based on work in his group. After studying the acid dependence of alkylation and finding that the rate change is negligible over physiologically relevant pH 7 and 8, they disregard the idea that the alkylation reaction is acid catalyzed.[5] Their models suggests that the basis for the stability of the duocarmycins is inherent in the vinylogous amide stabilization found in all the members of this family. Crystal structures of the duocarmycins, as well as many analogues show that these molecules are planar, a conformation that maximizes π orbital overlap.[6] Additional data from NMR structures reveal that when bound to DNA, the compounds are now shown to be bent, meaning that the vinylogous amide stabilization is lost.[7,8] Boger's model states that it is not only the reactivity of the electrophilic cyclopropane ring that drives the reaction, but also the loss of the vinylogous amide stabilization that occurs upon binding to DNA. Upon binding, if a nucleophile (N3 of adenine) is present, the alkylation event occurs. The stability of the

adduct is further reinforced because retroalkylation via spirocyclization is no longer favored because of this conformational change (Figure 2.2b).[9]

The target sequences of (+)-Duocarmycins A and SA are identical in their sequence preference with 5'-AAA-3' as the most preferred site (alkylated base underlined).[1] There is also a preference for A or T over G or C to the 5' side, and a weak preference of a purine over a pyrimidine at the 3' side. (+)-CC-1065 has a similar consensus sequence to the (+)-duocarmycins of 5'-(A/T)(A/T)AAA-3'. It has a longer 5 base pair binding site because of the additional benzopyrrole subunit. The unnatural enantiomers of duocarmycins also alkylate DNA, but typically ten times higher concentrations are necessary than the natural enantiomers to see comparable amounts of DNA cleavage. The consensus sequences are similar, but the binding orientations are reversed, due to the chirality of the reactive cyclopropyl moiety.

Analogues of CC-1065

Many analogues of CC-1065 have been synthesized in attempts to minimize the delayed toxicity effects while maximizing DNA alkylation efficiency. One of the early analogues synthesized in Dale Boger's group was 1,2,9,9a-tetrahydrocyclopropa-[c]-benz[e]indol-4-one (CBI).[10] Despite the rather significant structural change from CC-1065 to CBI (Figure 2.3), the alkylation properties of CBI were superior to that of CC-1065. CBI has proven to be more stable at biologically relevant pH and more potent as an alkylating agent than CC-1065. Because of its simplified structure, it is also more readily synthesized. Earlier work had also shown that the seco (ring open) agents had indistinguishable DNA alkylation selectivities, efficiencies and biological potencies as the cyclopropyl derivatives.[11] We chose to use the *seco*-CBI agents to conjugate to the

polyamide because of their easier accessibility and longer shelf lives as well as their increased reactivity to DNA. In addition we chose to examine both enantiomers of *seco*-CBI in order to determine if strand selective alkylation could be achieved by the enantiomer chosen.

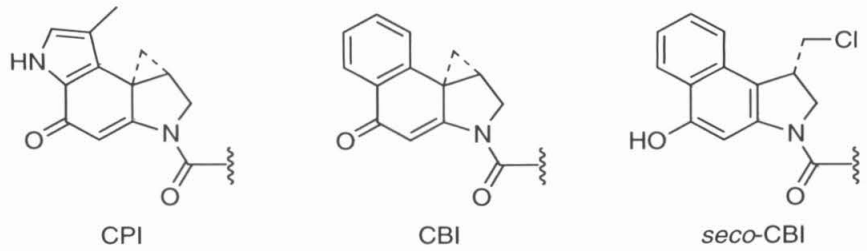


Figure 2.3. Analogues of CC-1065.

Selection of Hairpin Polyamide

The eight-ring hairpin polyamide of sequence composition ImImPyPy-(R)$^{H_2N}\gamma$-ImPyPyPy-β-Dp (**1**) (Figure 2.4) was designed as the parent polyamide of this series. This polyamide had been previously well characterized and targets the match site 5'-WGGWCW-3'. Polyamide **1** has been shown to target the sequence 5'-TGGTCA-3' with a K_a of 1.3×10^{10} M^{-1}.[12] We chose polyamide **1** because it has been well characterized and also because it targets a non-palindromic sequence. For the early members of this series of compounds, we wanted to be able to examine the efficiency of cleavage without the concerns of more than one binding orientation of the polyamide. The polyamide conjugates **2R** and **2S** are also shown in Figure 2.4. Glutaric acid links *seco*-CBI to the polyamide via the chiral turn.[13] The control compounds **3R** and **3S** (Figure 2.4) incorporate the *seco*-CBI subunit and a positive charge for solubility. The sequence specificity of the *seco*-CBI conjugates will be determined by comparing alkylation patterns and yields to the control compounds.

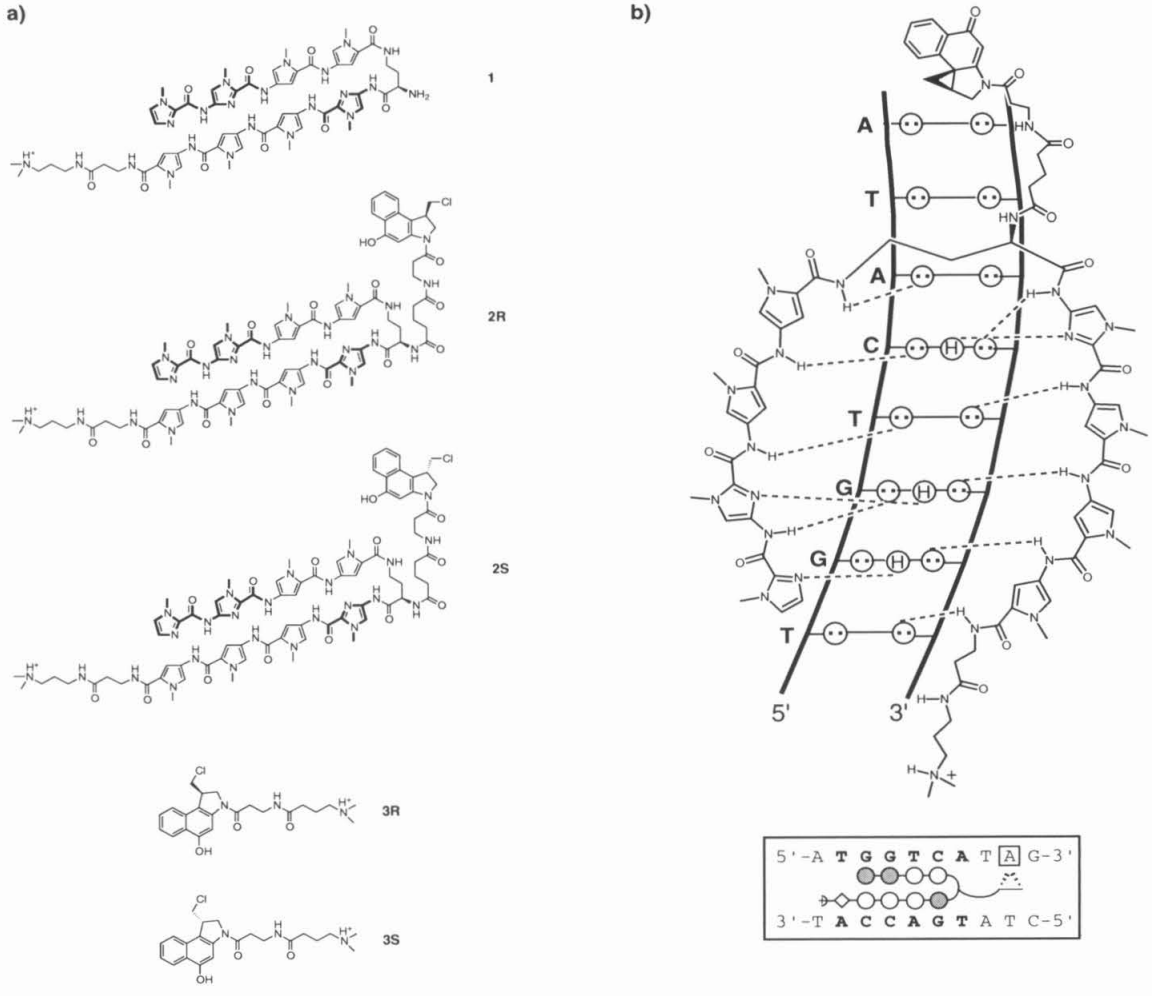


Figure 2.4. (a) Structures of Parent polyamide **1**, CBI-conjugates **2R** and **2S**, and unlinked control compounds **3R** and **3S**. (b) (top) Hydrogen bond model of the polyamide-DNA complex formed by the polyamide ImImPyPy-$\gamma^{(S\text{-}CBI)}$-ImPyPyPy-β-Dp (**2S**) bound to the minor groove of 5'-TGGTCA-3'. Circles with dots represent lone pairs of N3 purines and O2 of pyrimidines. Circles containing an H represent the N2 hydrogen of G. Putative hydrogen bonds are illustrated by dotted lines. (bottom) Binding model for polyamide ImImPyPy-$\gamma^{(S\text{-}CBI)}$-ImPyPyPy-β-Dp (**2S**) with a 5'-TGGTCA-3' site. Shaded and nonshaded circles denote imidazole (Im) and pyrrole (Py) rings, respectively. Diamonds and hatched triangles represent β-alanine (β) and (S)-CBI, respectively. (R)-2,4-diaminobutyric acid (γ) and dimethylaminopropylamine (Dp) are depicted as a curved line and a plus sign, respectively.

Synthesis of Boc-seco-CBI

Synthesis of Boc protected-*seco*-CBI **4** was performed according to the procedures outlined by Boger in 1995 (Figure 2.5).[14-16] Briefly, condensation of ammonia and 1,3-dihydroxynaphthalene with subsequent Boc protection of the amine and benzyl protection of the alcohol yields the protected naphthalene derivative. Treatment with NIS provides the iodonapthylamine derivative. Alkylation with allyl bromide provided a substrate for a favorable 5-exo-trig aryl radical-alkene cyclization to occur, using Bu_3SnH and TEMPO radical trap. Cleavage of the TEMPO trap intermediate occurs upon heating with activated Zn powder. Treatment with PPh_3/CCl_4 followed by deprotection of the benzyl ether gave the desired product **4** in racemic form.

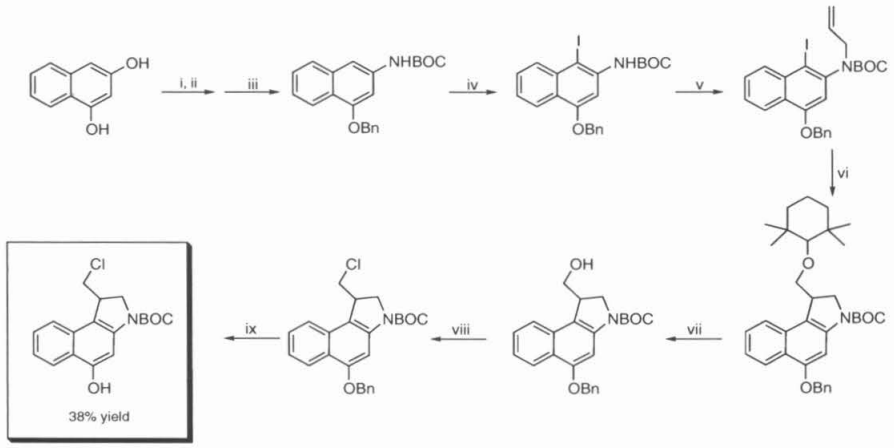


Figure 2.5. Synthesis of Boc-*seco*-CBI (**4**). (i) NH_3 (liquid); (ii) Boc anhydride; (iii) benzyl bromide; (iv) NIS; (v) allyl bromide, sodium hydride; (vi) TEMPO, Bu_3SnH; (vii) Zn/AcOH; (viii) PPh_3/CCl_4; (ix) HCO_2NH_4, Pd/C.

Chiral resolution of racemic *seco*-CBI **4** is accomplished through chiral HPLC methods also developed by Boger.[17] Using a semi-preparative column, one can purify 50 to 100 mg of material over several runs. For the coupling chemistry used for polyamide conjugate synthesis, it was found that the secondary amine of **4** (after deprotection) was difficult to couple to the polyamide. For this reason, we added a β-alanine linker which

couples well to the NHS ester of the polyamide. Synthesis of Boc-β-ala-*seco*-CBI **5** is illustrated in figure 2.6. Deprotection with HCl/ethyl acetate followed by addition of Boc-β-alanine, HBTU, and DIEA in DMF yields the desired analogue **5**.

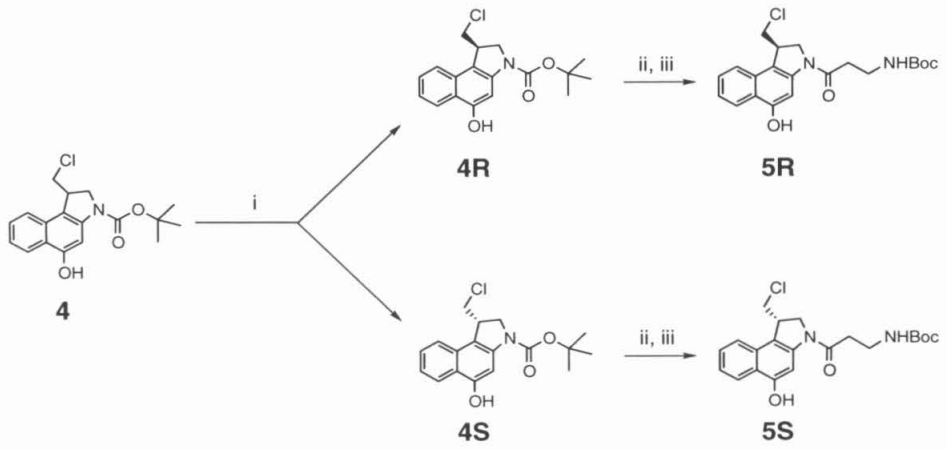


Figure 2.6. Synthesis of Boc-β-alanine-*seco*-CBI (**5R** and **5S**). (i) ChiralCel OD; (ii) 3M HCl/EtOAc, 30 min; (iii) Boc-β-Alanine, HBTU, DIEA, DMF.

Synthesis of Polyamide-seco-CBI Conjugates and Unlinked Analogues

The eight-ring hairpin polyamide **1** was prepared by manual solid-phase peptide synthesis. [18] After reverse phase HPLC purification, an NHS-activated glutaric acid linker was attatched to the α-amino group on the γ-turn to afford modified polyamide **1-DSG**.[13] After HPLC purification, the appropriate enantiomer of β-alanine linked *seco*-CBI was coupled using DCC/NHS activation to give the corresponding polyamide-*seco*-CBI conjugates **2R** and **2S** (Figure 2.7). The control *seco*-CBI analogues **3R** and **3S** with the same charge as the polyamide analogues were prepared by coupling the β-alanine derivatives **5R** and **5S** to N,N-dimethyl-γ-amino butyric acid. Purification by reverse phase HPLC yields compounds **3R** and **3S** (Figure 2.8).

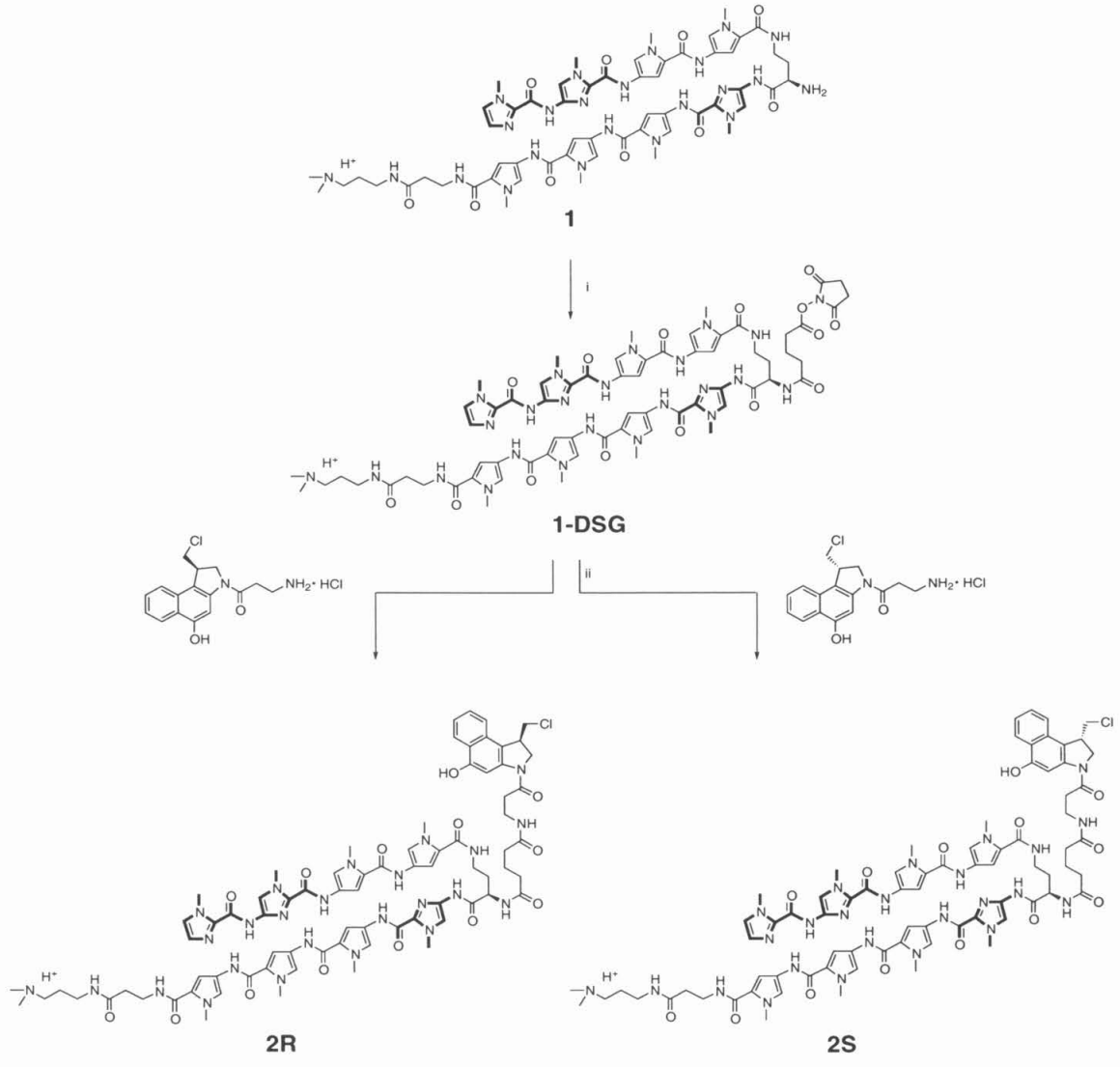


Figure 2.7. Synthetic scheme for preparation of *seco*-CBI-polyamide conjugates **2R** and **2S**. Polyamide **1** was prepared by already disclosed methods. (i) Disuccinimidyl Glutarate, DMF, 2 H (ii) amine of **5R** or **5S**, DCC, NHS, DIEA, DMF, 3H.

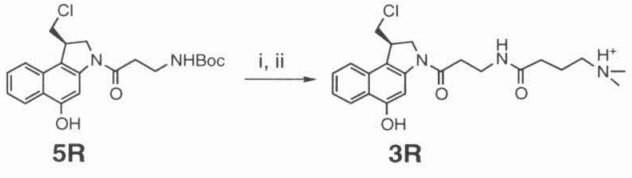


Figure 2.8. Synthesis of unlinked control compound **3R**. (i) 3M HCl/EtOAc, 30 min; (ii) 4-Dimethylamino butyric acid, HBTU, DIEA, DMF.

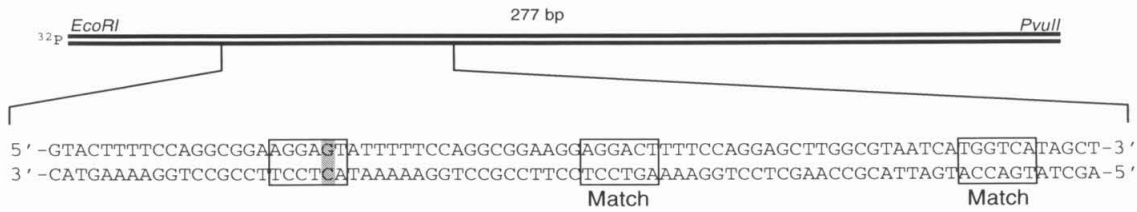


Figure 2.9. Illustration of the 277 base pair *EcoRI/HindIII* restriction fragment with the position of the sequence indicated. The binding sites for polyamide 1, 5'-AGGACT-3' and 5'-TGGTCA-3' are boxed. A single base pair mismatch site is also boxed, the mismatch shaded.

Thermally induced cleavage reactions

To examine the alkylation specificity and reaction yields of the polyamide-*seco*-CBI conjugates **2R** and **2S**, thermally induced cleavage assays were performed on a 277 base pair restriction fragment containing two match sites, 5'-AGGACT-3' and 5'-TGGTCA-3' and one mismatch site, 5'-AGGAGT-3' (mismatch underlined), with different A/T tracts flanking the 3' side (Figure 2.9). Polyamide **2R** at a concentration of 500 pM alkylates a single adenine on the 3'-labeled (bottom) strand two base pairs to the 3' side of the match site 5'-AGGACT-3' with a yield of 74.5% (Figure 2.10b). At 200 pM concentrations of polyamide **2R** a second cleavage site appears on the same strand at a single adenine proximal to the mismatch site 5'-AGGAGT-3'. At 1nM concentrations of **2R** and 12 hours of equilibration near quantitative cleavage (96%) of the intact DNA is observed (Figure 2.10b). Remarkably, the reaction appears to be strand specific. No cleavage appears on the 5'-labeled (top) strand (Figure 2.10a).

In constrast, polyamide **2S** at 50 pM concentrations affords one major cleavage site on the 5'-labeled (top) strand corresponding to the adenine two base pairs to the 3' side of the polyamide match site 5'-TGGTCA-3'. At 1 nM concentrations of **2S** cleavage yields of 42% are observed (Figure 2.10a). On the 3' labeled (bottom) strand, only minor

Figure 2.10. Thermally induced strand cleavage on the 5'-end labeled and 3'-end labeled 277 base pair restriction fragment by ImImPyPy-$\gamma^{(R\text{-}seco\text{-}CBI)}$-ImPyPyPy-β-Dp (**2R**) and ImImPyPy-$\gamma^{(S\text{-}seco\text{-}CBI)}$-ImPyPyPy-β-Dp (**2S**). Storage phosphor autoradiograms of 8% denaturing polyacrylamide gels used to separate the fragments generated by heat induced DNA cleavage at alkylation sites. All lanes contain 10 kcpm of either 5' or 3' radiolabeled DNA. Each reaction was equilibrated in TE, pH 7.5 at 37 °C for 12 H. The unbound polyamide was removed by precipitation, and then strand cleavage was induced by heating at 95 °C for 30 min. (a) 5'-^{32}P-end labeled restriction fragment. (b) 3'-^{32}P-end labeled restriction fragment. (a-b) lanes 1 and 14, intact DNA; lanes 2-11, 15-24, 1 pM, 2 pM, 5 pM, 10 pM, 20 pM, 50 pM, 100 pM, 200 pM, 500 pM, 1 nM respectively of the corresponding polyamide; lanes 12 and 25, A-specific reaction; lanes 13 and 26, G-specific reaction. (a-b) Match sites 5'-TGGTCA-3' and 5'-AGGACT-3' and single base pair mismatch site 5'-AGGAGT-3' are indicated in bold on the sequence, with arrows indicating cleavage bands.

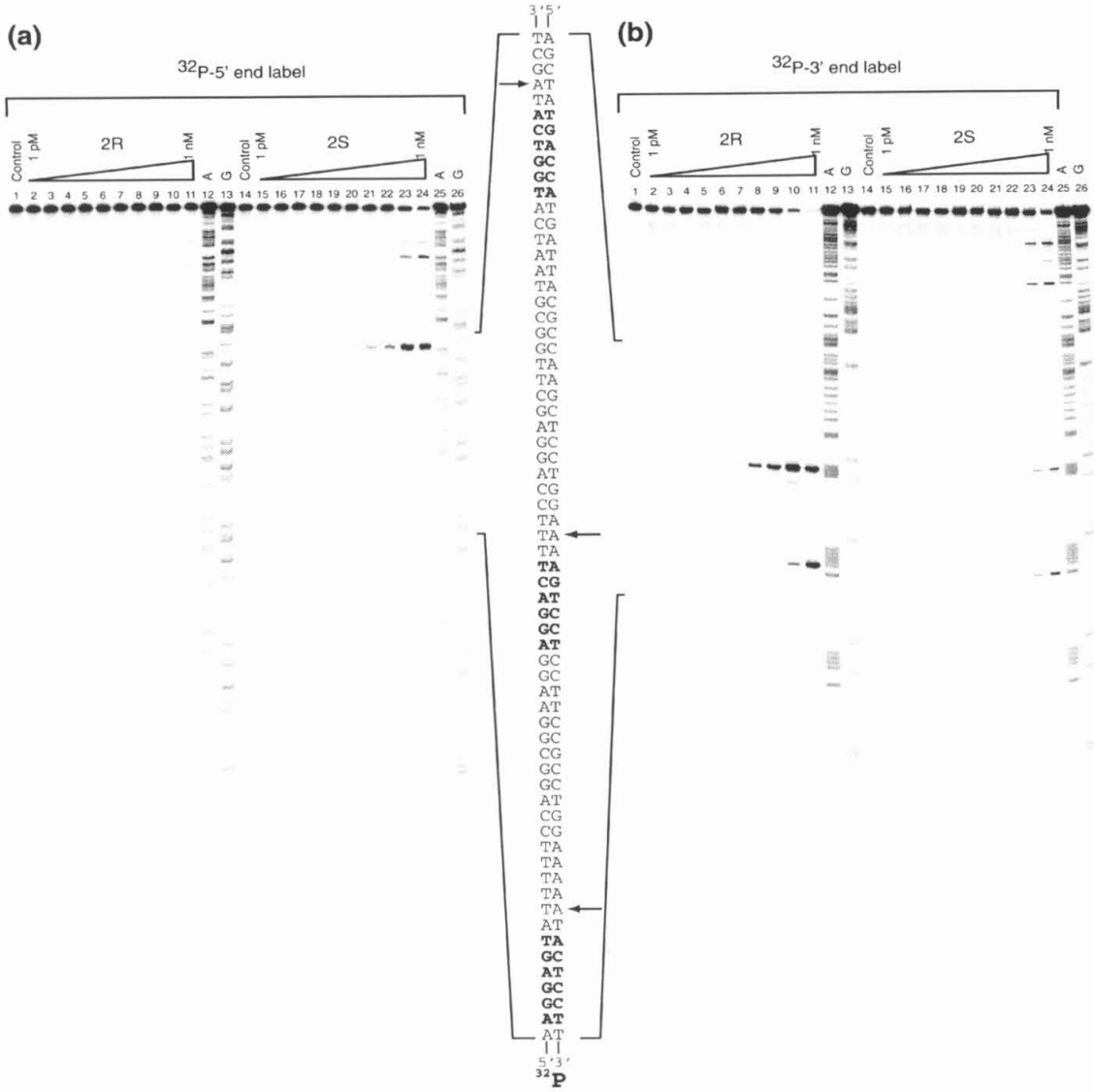
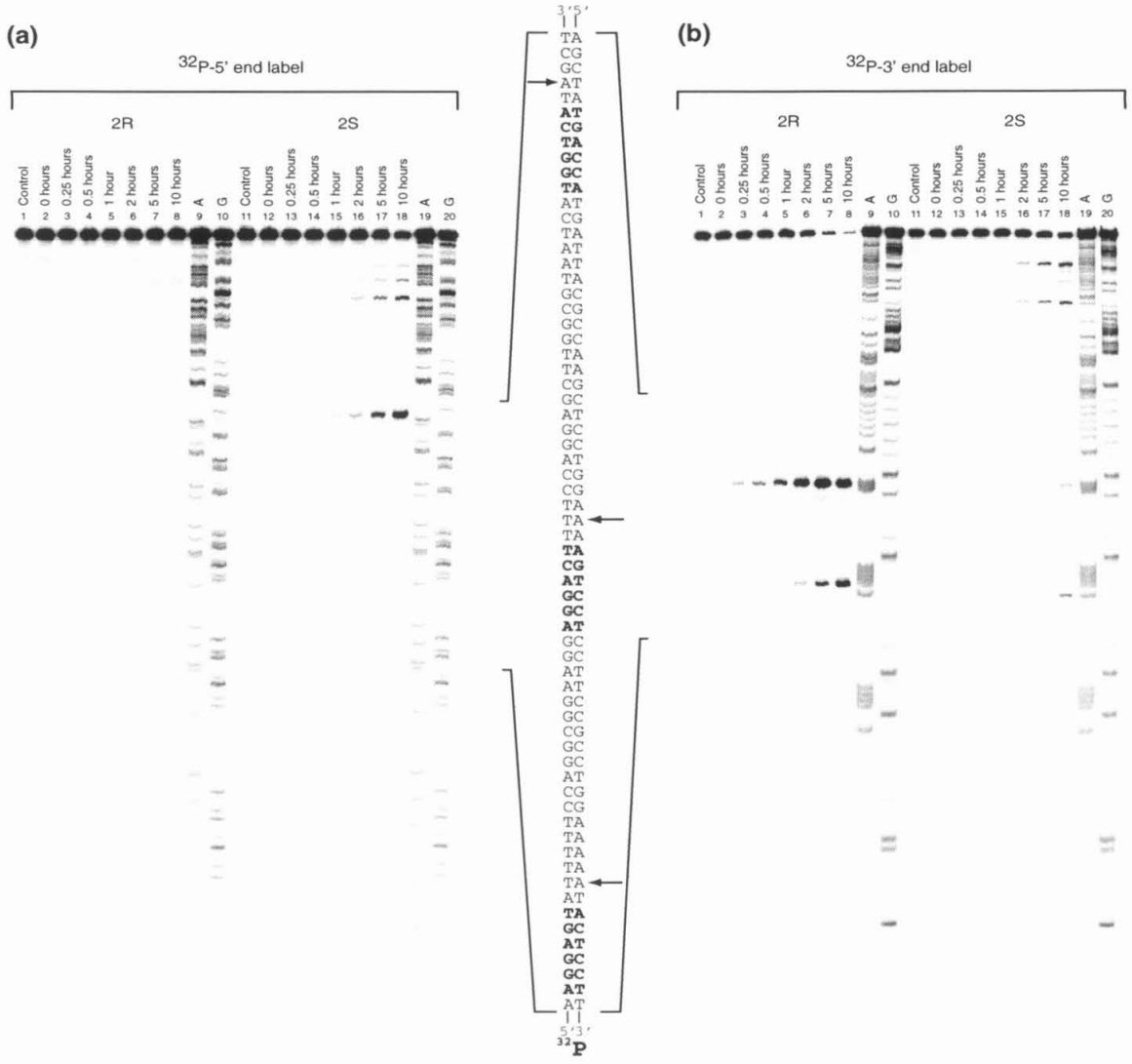

(a)
32P-5' end label
Control
1 pM
2R
1 nM
A
G
Control
1 pM
2S
1 nM
A
G
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26
3'
5'
TA
CG
GC
AT
TA
AT
CG
TA
GC
GC
TA
AT
CG
TA
AT
AT
TA
GC
CG
GC
GC
TA
TA
CG
GC
AT
GC
GC
AT
CG
CG
TA
TA
TA
TA
CG
GC
AT
GC
GC
AT
GC
GC
AT
AT
GC
GC
CG
GC
GC
AT
CG
CG
TA
TA
TA
TA
TA
AT
TA
GC
AT
GC
GC
AT
AT
5'
3'
32P
(b)
32P-3' end label
Control
1 pM
2R
1 nM
A
G
Control
1 pM
2S
1 nM
A
G
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26

Figure 2.11. Time course experiment of thermally induced strand cleavage on the 5'-end labeled and 3'-end labeled 277 base pair restriction fragment by ImImPyPy-$\gamma^{(R\text{-}seco\text{-}CBI)}$-ImPyPyPy-β-Dp (**2R**) and ImImPyPy-$\gamma^{(S\text{-}seco\text{-}CBI)}$-ImPyPyPy-β-Dp (**2S**). Storage phosphor autoradiograms of 8% denaturing polyacrylamide gels used to separate the fragments generated by heat induced DNA cleavage at alkylation sites. All lanes contain 10 kcpm of either 5' or 3' radiolabeled DNA. Each reaction was equilibrated in TE, pH 7.5 at 37 °C from 0 to 10 hours. The unbound polyamide was removed by precipitation, and then strand cleavage was induced by heating at 95 °C for 30 min. (a) 5'-^{32}P-end labeled restriction fragment. (b) 3'-^{32}P-end labeled restriction fragment. (a-b) lanes 1 and 11, intact DNA; lanes 2-8, 1nM of corresponding polyamide, equilibrations for 0, 15 minutes, 30 minutes, 1 hour, 2 hours, 5 hours, and 10 hours, respectively; lanes 9 and 19, A-specific reaction; lanes 10 and 20, G-specific reaction. (a-b) Match sites 5'-TGGTCA-3' and 5'-AGGACT-3' and single base pair mismatch site 5'-AGGAGT-3' are indicated in bold on the sequence, with arrows indicating cleavage bands.


(a)
32P-5' end label
2R
2S
Control
0 hours
0.25 hours
0.5 hours
1 hour
2 hours
5 hours
10 hours
A
G
(b)
32P-3' end label
3'5'
5'3'
32P

alkylation and cleavage is observed at four sites in low yield (Figure 2.10b). The two sites which can be assigned are proximal to a match 5'-AGGACT-3' and mismatch site 5'-AGGAGT-3', respectively.

Time Course Experiments

To examine the time dependence of alkylation, reactions consisting of polyamides **2R** and **2S** at 1 nM concentrations were analyzed throughout a period of 10 hours. For the 5'-labeled strand, alkylation by **2S** was first detected at 1 hour of incubation at 37 °C. After 10 hours, 24% of the DNA was cleaved proximal to the match site, 5'-TGGTCA-3'. **2R** shows no alkylation on this strand (Figure 2.11). However, **2R** shows alkylation on the 3'-labeled strand after 15 minutes, and at 5 hours, has 76% yield proximal to the match site 5'-AGGACT-3', and 13% at the mismatch site 5'-AGGAGT-3'. After 10 hours, **2S** reveals minor cleavage products on the 3'-labeled strand, as described above for the titration experiments, each ranging from 1.5%-17% yield (Figure 2.11).

Comparison of Conjugates to non-linked analogues

In order to study the ability of polyamides to direct the reactivity of CBI analogues, alkylation by the unlinked analogues were analyzed as controls (Figure 2.12). Analogue **3R** shows virtually no cleavage products at 1 μM concentrations, three orders of magnitude higher concentrations than that used in the conjugate studies. At 10 μM concentrations, **3R** shows several cleavage products, but with yields only between 1.1% and 5.1% on either strand. **3S** is shown to be a more efficient alkylating agent than **3R**. On the 5'-labeled strand, at 1 μM **3S**, the majority of the cleavage products are at 1% yield or lower. On the 3'-labeled strand, the efficiency ranges from less than 1% to 7.7% at the same concentrations. At 10 μM **3S**, the experiments show quantitative cleavage of

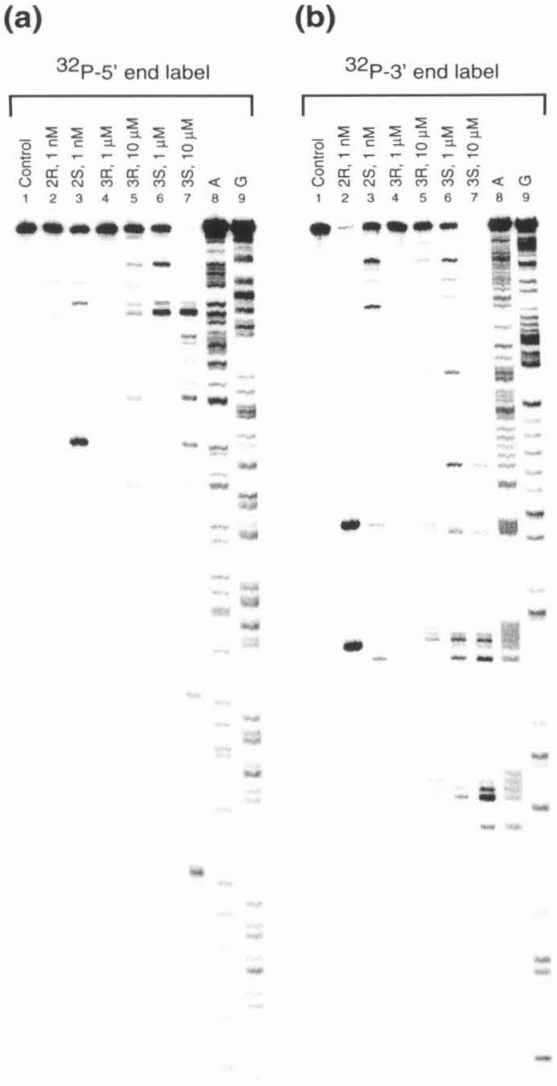
the intact DNA for both strands, and the efficiency of alkylation on all sites increases accordingly. The yields range from 2%-25% and 2%-35% on the 3' and 5'-labeled strands, respectively. The consensus sequence for both enantiomers of CBI is an adenine for alkylation, a >95% preference for an A•T base pair to the 5' side of the adduct, and a 66% preference for a purine to the 3' side. The alkylation pattern observed for both **3R** and **3S** shows that the AT tracts are particularly reactive sites, but the selectivity is difficult to predict (Figure 2.13).

Discussion

Strand Selective Cleavage

One of the striking differences between polyamide-*seco*-CBI conjugates **2R** and **2S** is the strand selectivity of alkylation shown by the respective diastereomers (Figure 8). This is the result of combining the DNA binding properties of the polyamide and the opposite strand alkylation preference for each mirror image of CBI in the minor groove of DNA. The data supports a model wherein the hairpin polyamide binds both match and mismatch sites and directs the chiral CBI moiety to A/T tracts proximal to the bound sites. The electrophilic cyclopropyl carbon of the R and S-CBI reacts at different rates with the N3 of adenine on opposite strands on the floor of the minor groove. The specificity of reaction at specific adenines represent the ratios of two unimolecular rate constants: the dissociation rate of the polyamide from the minor groove of DNA (k_{OFF}) and the alkylation rate of CBI at N3 of adenine in the minor groove of DNA (k_{ALK}). When the polyamide binds to a match site, k_{OFF} is expected to be slow relative to mismatch sites. Similarly, k_{ALK} will be faster for sites which accommodate the orientation

Figure 2.12. Thermally induced strand cleavage on the 5'-end labeled and 3'-end labeled 277 base pair restriction fragment by ImImPyPy-$\gamma^{(R\text{-}seco\text{-}CBI)}$-ImPyPyPy-β-Dp (**2R**), ImImPyPy-$\gamma^{(S\text{-}seco\text{-}CBI)}$-ImPyPyPy-β-Dp (**2S**), (R)-*seco*-CBI-β-dimethyl-γ (**3R**), (S)-*seco*-CBI-β-dimethyl-γ (**3S**). Storage phosphor autoradiograms of 8% denaturing polyacrylamide gels used to separate the fragments generated by heat induced DNA cleavage at alkylation sites. All lanes contain 10 kcpm of either 5' or 3' radiolabeled DNA. Each reaction was equilibrated in TE, pH 7.5 at 37 °C for 12 H. The unbound polyamide or agent was removed by precipitation, and then strand cleavage was induced by heating at 95 °C for 30 min. (a) 5'-^{32}P-end labeled restriction fragment. (b) 3'-^{32}P-end labeled restriction fragment. (a-b) Lane 1, intact DNA; lane 2, **2R**, 1nM; lane 3 **2S**, 1nM; lanes 4-5, **3R**, 1 μM and 10 μM, respectively; lanes 6-7, **3S**, 1 μM and 10 μM, respectively; lane 8, A-specific reaction; lane 9, G-specific reaction.


(a)
32P-5' end label
1 Control
2 2R, 1 nM
3 2S, 1 nM
4 3R, 1 µM
5 3R, 10 µM
6 3S, 1 µM
7 3S, 10 µM
8 A
9 G
(b)
32P-3' end label
1 Control
2 2R, 1 nM
3 2S, 1 nM
4 3R, 1 µM
5 3R, 10 µM
6 3S, 1 µM
7 3S, 10 µM
8 A
9 G

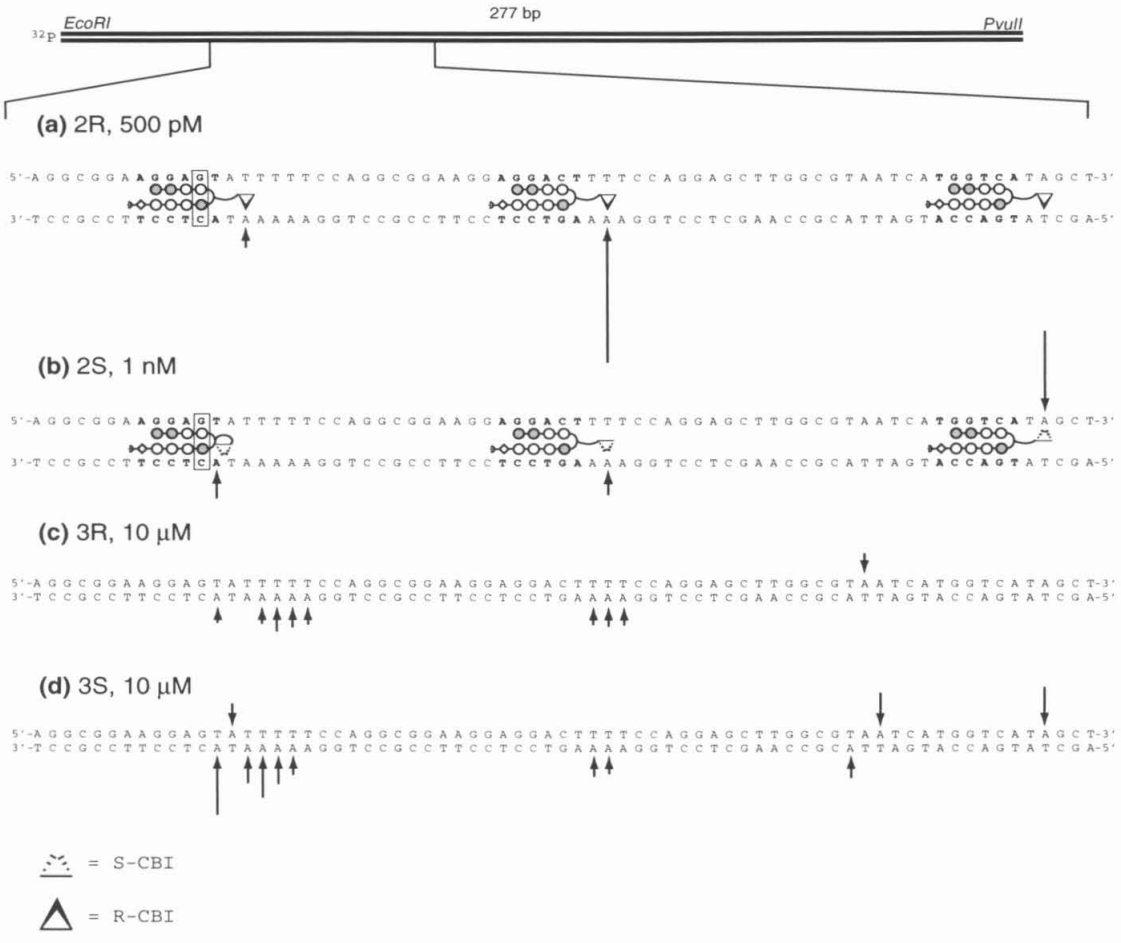


Figure 2.13. Illustration of the 277 bp restriction fragment with the position of the sequence indicated. Cleavage patterns are shown for (a) ImImPyPy-$\gamma^{(R\text{-}seco\text{-}CBI)}$-ImPyPyPy-β-Dp (**2R**), 500 pM; (b) ImImPyPy-$\gamma^{(S\text{-}seco\text{-}CBI)}$-ImPyPyPy-β-Dp (**2S**), 1 nM; (c) (R)-*seco*-CBI-β-dimethyl-γ (**3R**), 10 μM; and (d) (S)-*seco*-CBI-β-dimethyl-γ (**3S**), 10 μM. Match sites 5'-TGGTCA-3' and 5'-AGGACT-3' and single base pair mismatch site 5'-AGGAGT-3' are indicated in bold on the sequence, with arrows indicating cleavage bands. The single base pair mismatch is indicated by a rectangle. The polyamide is colored as in Figure 2.4b, with the bold triangle representing (R)-CBI.

and sequence preferences of CBI as in pure A/T tracts, and slower for mixed sequences. These ratios of k_{OFF}/k_{ALK} at different DNA sites give the variation in alkylation yields seen in these experiments. On the 5'-labeled (top) strand, polyamide **2S** has one major cleavage site (Figure 2.10a and 2.11a) that is due to the polyamide binding the match site 5'-TGGTCA-3' (slow k_{OFF}) and an adenine in the appropriate position proximal to the match site for alkylation (fast k_{ALK}). Likewise, on the 3'-labeled (bottom) strand, polyamide **2R** has a major cleavage site (Figure 2.10b and 2.11b) due to the polyamide binding the match site 5'-AGGACT-3' (slow k_{OFF}) and an adenine on the opposite strand for alkylation (fast k_{ALK}). For the two DNA match sites bound by each diastereomeric polyamide-*seco*-CBI conjugate (**2R** and **2S**), k_{OFF} rates are expected to be similar. The variation in alkylation yields in these two instances most likely results from k_{ALK} being faster for the sequence proximal to the match site for **2R** than for that of **2S**.

Similar comparisons can be made for alkylation seen at mismatch sites. On the 3'-labeled (bottom) strand, polyamide **2R** has a minor cleavage site (Figure 2.10b and 2.11b) due to the polyamide binding the single base pair mismatch site 5'-AGGAGT-3' (fast k_{OFF}). But, because of favorable sequence contexts k_{ALK} is also fast. The alkylation at this site results from k_{ALK} being competitive with k_{OFF}, trapping that particular binding event. Also on the 3'-labeled (bottom) strand, polyamide **2S** has two minor cleavage sites that can be assigned (Figure 2.10b and 2.11b). For the alkylation seen adjacent to the match site 5'-AGGACT-3', k_{OFF} is expected to be slow. In this case, k_{ALK} is also expected to be slow because the orientation preferences of S-*seco*-CBI are not optimal at this site. However, the binding event can be trapped because k_{ALK} may be competitive with k_{OFF} in this instance. **2S** also affords an alkylation product next to the single base pair mismatch

site 5'-AGGAGT-3'. While k_{OFF} is expected to be faster in this case than for that of a match site, it can still be trapped for situations where k_{ALK} is faster than k_{OFF}.

It has been shown for CC-1065 and other analogues that the S, or (+)-enantiomers, are 10 times more reactive than the R, or (-)-enantiomers.[4,19,20] The data shown for the control compounds **3R** and **3S** support this conclusion (Figure 2.12). It is interesting, then, that polyamide conjugate **2R** shows higher alkylation efficiency and faster rates of alkylation than **2S**. Work by Lukhtanov and coworkers have shown that when an oligonucleotide-cyclopropapyrroloindole conjugate is hybridized to a variety of complementary oligonucleotide hairpins, rates of reaction can be quite rapid, and yields can be quantitative.[21] But the results are sequence dependent. In the case of the oligonucleotide conjugates, both enantiomers show fast and efficient alkylation in pure A/T tracts ($t_{1/2}$ as fast as 2 minutes), while mixed sequences seem to have slower and more unpredictable rates of reactivity. In our studies, the pure A/T tracts flanking the polyamide match sites differ and we may simply be observing sequence dependent differences in rates of reaction as a result of different adenine reactivities in various sequence contexts.

Comparison with hairpin polyamide Duocarmycin A conjugates

Previous work by Sugiyama and coworkers has established that Duocarmycin A (Duo), an analogue of CC-1065, will alkylate a guanine preferentially when bound as a heterodimer with distamycin.[22] Sugiyama and coworkers have recently described pyrrole-imidazole hairpin polyamide conjugates which incorporate the natural enantiomer of Duo at the C-terminus.[23] These molecules demonstrate the same altered reactivity at guanine, and target purines proximal to a hairpin polyamide binding site. Our choice of

using *both enantiomers* of *seco*-CBI was driven by the orientation preferences of CC-1065 necessary for alkylation. These orientation preferences suggested that it might be possible to target either strand of the DNA duplex, based on the binding orientation of *seco*-CBI at that site. We have shown that the same hairpin polyamide can be used to target opposite strands of the DNA depending on the enantiomer of *seco*-CBI we choose. With the 6 ring hairpin polyamide-Duo conjugates, the highest yield Sugiyama and coworkers report is 7.4% cleavage at 800 nM polyamide for 7 days at room temperature.[23] Despite the rather slow reaction kinetics, this compound shows very good specificity for its match site. No other cleavage sites are seen for this compound on the fragment analyzed. Our 8 ring hairpin polyamide-*seco*-CBI conjugate shows higher yields of alkylation at faster rates and lower concentrations of polyamide, 42% (at a single site) for **2S** at 1 nM after 12 hours, and 74.5% (at a single site) for **2R** at 500 pM after 12 hours, both at 37°C. But we do see contributions from mismatch alkylation sites, 7.4% for **2R** at 500 pM after 12 hours at 37°C. Because the two systems are very distinct, it is difficult to interpret the differences in the results. It is unclear how differences in hairpin polyamide (6 ring versus 8 ring), place of attachment (C-terminus versus turn), linker length (amide versus longer methylene), and alkylation conditions (time, temperature, ligand concentration, DNA concentration) will affect the overall reaction. It is worth noting that both sets of results show alkylation directed by hairpin polyamides in a sequence specific fashion. Differences in yields and rates of reaction can be a consequence of many factors, which will require more studies for detailed understanding.

Conclusion.

Hairpin polyamide-CBI conjugates have been shown to efficiently and selectively alkylate a single adenine adjacent to a polyamide match site. Because of the high efficiency of alkylation, these molecules should be useful in the design of reagents that target a single gene. It has already been shown that triplex forming oligonucleotide-psoralen and nitrogen mustard conjugates form covalent adducts in the major groove on the coding strand to inhibit elongation of transcription.[24] It remains to be seen whether this class of polyamide-CBI conjugates which react in the minor groove will be useful for functional genomics.

Experimental

Materials^{1}H NMR spectra were recorded on a General Electric-QE NMR spectrometer at 300 MHz and a Varian Inova NMR spectrometer at 500 MHz with chemical shifts reported in parts per million relative to residual solvent. UV spectra were measured in water on a Hewlett-Packard Model 8452A diode array spectrophotometer. Optical rotations were recorded on a JASCO Dip 1000 digital polarimeter. Matrix-assisted, laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF) was performed at the Protein and Peptide Microanalytical Facility at the California Institute of Technology. High resolution mass spectometry was performed at the mass spectrometry facility at University of California at Los Angeles. Preparatory reversed phase HPLC was performed on a Beckman HPLC with a Waters DeltaPak 25 × 100 mm, 300 Å C18 column equipped with a guard, 0.1% (wt/v) TFA, 0.25% acetonitrile/min.

Synthesis of CBI-polyamide conjugates

ImImPyPy-(R)^{H_2N}γ-ImPyPyPy-β-Dp (1). ImImPyPy-(R)^{H_2N}γ-ImPyPyPy-β-Pam resin was synthesized in a stepwise fashion by Boc-chemistry manual solid phase protocols.[18] A sample of resin was treated with neat (dimethylamino)-propylamine (2 ml), heated (55 °C, 24 hours) and purified by reversed phase HPLC. ImImPyPy-(R)^{H_2N}γ-ImPyPyPy-β-Dp was recovered as a white powder upon lyophilization of the appropriate fraction (14.3 mg, 11.6 mmoles, 4.8% recovery). UV (H_2O) λ_{max} (ε), 312 nm, (66, 600); ^{1}H NMR (500 MHz, DMSO-d_6): δ=11.009 (s, 1H), 10.325 (s, 1H), 10.073 (s, 1H), 9.947 (s, 1H), 9.936 (s, 1H), 9.870 (s, 1H), 9.07 (s, 1H), 9.23 (br s, 1H, CF_3COOH), 8.334 (s, 3H), 8.173 (t, 1H, J=6 Hz), 8.040 (t, 1H, J=6 Hz), 8.012 (t, 1H, J=6 Hz), 7.567 (s, 1H), 7.530 (s, 1H), 7.457 (s, 1H), 7.270 (s, 1H), 7.256 (s, 1H), 7.200 (s, 1H), 7.177 (s, 1H), 7.164 (s, 1H), 7.155 (s, 1H), 7.142 (s, 1H), 7.071 (m, 2H), 6.964 (s, 1H), 6.876 (s, 1H), 4.002 (m, 6H), 3.978 (s, 3H), 3.853 (s, 3H), 3.842 (s, 3H), 3.833 (s, 3H), 3.807 (s, 3H), 3.792 (s, 3H), 3.375 (q, 2H, *J*=5.5 Hz), 3.294 (m, 2H, *J*=5.5 Hz), 3.106 (q, 2H, *J*=6.5 Hz), 2.997 (m, 3H), 2.735 (s, 3H), 2.725 (s, 3H), 2.343 (t, 2H, *J*=7 Hz), 1.738 (m, 2H), 1.633 (m, 2H). MALDI-TOF-MS (monoisotopic) [M+H] 1238.83 (calculated 1238.58 for C57H72N23O10.)

ImImPyPy-(R)$^{Glu\text{-}NHS}$γ-ImPyPyPy-β-Dp (1-DSG). To a solution of disuccinimidyl glutarate (41.9 mg, 120 μmoles) in 2.5 ml DMF was added 100 ml of a 14.3 mM solution of **1** (15.9 mg, 12.8 μmoles) in DMF (800 μl) and DIEA (100 μl). 100 ml of the solution was added every 15 minutes while stirring. Following the completion of the addition of **1**, the reaction was stirred for 2 hours. The reaction was diluted with 0.1% TFA (15 ml) and the reaction was purified by reversed phase HPLC. ImImPyPy-(R)$^{Glu\text{-}NHS}$γ-

ImPyPyPy-β-Dp was recovered as a white powder upon lyophilization of the appropriate fraction (8.8 mg, 6.1 μmoles, 47.3% recovery). UV (H_2O) λ_{max} (ε), 312nm, (66, 600); ^1H NMR (500 MHz, DMSO-d_6, 25°C): δ=10.320 (s, 1H), 10.242 (s, 1H), 10.097 (s, 1H), 9.930 (s, 2H), 9.872 (s, 1H), 9.740 (s, 1H), 9.23 (br s, 1H, CF_3COOH), 8.183 (d, 1H, *J*=8 Hz), 8.034 (t, 1H, *J*=5.5 Hz), 8.010 (t, 1H, *J*=5.5 Hz), 7.946 (t, 1H, *J*=5.3 Hz), 7.563 (s, 1H), 7.454 (s, 1H), 7.272 (s, 1H), 7.263 (s, 1H), 7.215 (s, 1H), 7.181 (s, 1H), 7.144 (m, 4H), 7.080 (s, 1H), 7.061 (s, 1H), 6.887 (s, 1H), 6.870 (s, 1H), 4.001 (s, 6H), 3.957 (s, 3H), 3.848 (s, 3H), 3.842 (s, 3H), 3.833 (s, 3H), 3.793 (s, 3H), 3.790 (s, 3H), 3.7-3.9 (br, m, 3H), 3.378 (q, 2H, *J*=5.5 Hz), 3.228 (m, 2H), 3.107 (q, 2H, *J*=6 Hz), 2.998 (m, 2H), 2.741 (s, 3H), 2.731 (s, 3H), 2.345 (t, 2H, *J*=7 Hz), 2.228 (m, 2H), 1.728 (m, 4H). MALDI-TOF-MS (monoisotopic) [M+H] 1449.63 (calculated 1449.62 for $C_{66}H_{81}N_{24}O_{15}$).

ImImPyPy-(R)$^{(R)\text{-CBI}}$γ-ImPyPyPy-β-Dp (2R). To a solution of **6** (6.9 mg, 4.75 μmoles) in dry DMF was added 47.5 μl of a 1μM solution of DCC (10 equiv.) and 9.5 μl of a 0.5 M solution of N-hydroxysuccinimide (1 equiv.). The solution was stirred for 2 hours. Separately, a solution of **5R** (2 mg, 1 equiv.) was deprotected with 3M HCl/ethyl acetate (5 ml) for 30 minutes under argon. The ethyl acetate was then removed by vacuum and then coevaporated twice from dichloromethane to yield the amine of **5R**. The gray solid was then dissolved in 50 ml of dry DMF and then added to the polyamide solution. DIEA (8 ml, 10 equiv.) was then added and the reaction was stirred for 3 hours under argon. Upon completion, the reaction was diluted with 0.1% TFA (2 ml) and the reaction was purified by reversed phase HPLC. ImImPyPy-(R)$^{(R)\text{-CBI}}$γ-ImPyPyPy-β-Dp was recovered as a white powder upon lyophilization of the appropriate fraction (1.4 mg, 17.8% recovery). UV (H_2O) λ_{max} (ε), 314 nm, (73, 854); ^1H NMR (500 MHz, DMSO-d_6,

25°C): δ=10.342 (s, 1H), 10.316 (s, 1H), 10.248 (s, 1H), 10.096 (s, 1H), 9.928 (s, 2H), 9.869 (s, 1H), 9.691 (s, 1H), 9.17 (br s, 1H, CF_3COOH), 8.172 (d, 1H, *J*=7.5 Hz), 8.063 (d, 1H, *J*=8Hz) 8.029 (t, 1H, *J*=6 Hz), 8.008 (t, 1H, *J*=5.5 Hz), 7.962 (m, 2H), 7.894 (t, 1H, J= 5.5 Hz), 7.747 (d, 1H, *J*=8 Hz), 7.556 (s, 1H), 7.464 (t, 1H, *J*=7.5 Hz), 7.449 (s, 2H), 7.299 (t, 1H, J=7), 7.267 (s, 1H), 7.254 (s, 1H), 7.206 (s, 1H), 7.172 (s, 1H), 7.145 (m, 3H), 7.059 (s, 2H), 6.892 (s, 1H), 6.870 (s, 1H), 4.526 (q, 1H, *J*=7 Hz), 4.287 (t, 1H, *J*=10.5 Hz), 4.111 (m, 2H), 3.998 (s, 6H), 3.948 (s, 3H), 3.844 (s, 3H), 3.837 (s, 3H), 3.831 (s, 3H), 3.791 (s, 3H), 3.787 (s, 3H), 3.106 (q, 2H, *J*=6.5 Hz), 3.004 (m, 4H), 2.739 (s, 3H), 2.729 (s, 3H), 2.343 (t, 2H, *J*=7.3 Hz), 2.181 (t, 2H, *J*=7 Hz), 2.098 (t, 2H, *J*=7.8 Hz), 1.981 (m, 2H), 1.741 (m, 4H), 1.631 (m, 2H), 1.540 (m, 1H). MALDI-TOF-MS (monoisotopic) [M+H] 1638.72 (calculated 1638.69 for $C_{78}H_{93}N_{25}O_{14}$).

ImImPyPy-(R)$^{(S)\text{-}CBI}$γ-ImPyPyPy-β-Dp (2S). ImImPyPy-(R)$^{(S)\text{-}CBI}$g-ImPyPyPy-b-Dp was prepared from **6** as described for **2R**. (1.4 mg, 30.4% recovery). UV (H_2O) λ_{max} (ε), 314nm, (73, 854); ^{1}H NMR (DMSO-d_6): ^{1}H NMR (500 MHz, DMSO-d_6, 25°C): δ=10.343 (s, 1H), 10.316 (s, 1H), 10.248 (s, 1H), 10.095 (s, 1H), 9.929 (s, 2H), 9.870 (s, 1H), 9.689 (s, 1H), 9.19 (br s, 1H, CF_3COOH), 8.173 (d, 1H, *J*=7.5 Hz), 8.064 (d, 1H, *J*=8 Hz), 8.029 (m, 2H), d=7.963 (m, 2H, C4-H), 7.895 (t, 1H, *J*=5.5 Hz), 7.747 (d, 1H, *J*=8 Hz), 7.556 (s, 1H), 7.464 (t, 1H, *J*=8 Hz), 7.451 (s, 1H), 7.445 (s, 1H), 7.299 (t, 1H, *J*=7.8 Hz), 7.266 (s, 1H), 7.255 (s, 1H), 7.206 (s, 1H), 7.175 (s, 1H), 7.145 (m, 3H), 7.060 (s, 2H), 6.891 (s, 1H), 6.870 (s, 1H), 4.525 (q, 1H, *J*=7 Hz), 4.285 (t, 1H, *J*=10.5 Hz), 4.111 (d, 2H, *J*=8.5 Hz), 3.998 (s, 6H), 3.951 (s, 3H), 3.844 (s, 3H), 3.837 (s, 3H), 3.831 (s, 3H), 3.792 (s, 3H), 3.787 (s, 3H), 3.105 (q, 2H, *J*=6 Hz), 3.000 (m, 4H), 2.737 (s, 3H), 2.729 (s, 3H), 2.343 (t, 2H, *J*=6.8 Hz), 2.180 (t, 2H, *J*=8 Hz), 2.099 (t, 2H, *J*=8

Hz), *1.997* (m, *2H*), 1.727 (m, 2H), 1.631 (m, 2H), 1.540 (m, 1H). MALDI-TOF-MS (monoisotopic) [M+H] 1638.71 (calculated 1638.69 for $C_{78}H_{93}N_{25}O_{14}$).

3-(*tert*-Butoxycarbonyl)-1-(chloromethyl)-5-hydroxy-1,2-dihydro-3*H*-benz[e]indole (*seco*-CBI-BOC, 4) was synthesized and enantiomers separated by already published protocols.[14-16]

***Seco*-CBI-β-alanine-BOC (5R and 5S). 4** (85 mg, .255 mmoles) was deprotected in 3M HCl/ethyl acetate (10 ml) for 30 minutes under argon. After the ethyl acetate was removed by evaporation, and the deprotected *seco*-CBI was coevaporated twice from dichloromethane. BOC-b-alanine (96.4 mg, 2 equiv.) and EDC (293.3 mg, 6 equiv.) were added with DMF (5 ml). The solution was stirred overnight under argon. After the reaction was complete, 15 ml of water was added, and the reaction was extracted 6 times with ethyl ether. The ether was washed with brine, and dried with Na_2SO_4, and purified by flash chromatography (5% methanol/dichloromethane) to yield **5** as an off-white powder (90 mg, 86%).

(R)-Seco-CBI-β-alanine-BOC (**5R**): ^{1}H NMR (300 MHz, $CDCl_3$, 25°C): δ=9.37 (s, 1H), 8.27 (m, 2H), 7.62 (d, 1H, *J*=8.1 Hz), 7.52 (t, 1H, *J*=6.6 Hz), 7.38 (t, 1H, *J*=7.5 Hz), 5.55 (br t, 1H), 4.19 (d, 1H, J=10.2 Hz), 4.06 (t, 1H, J=9.9 Hz), 3.93 (m, 2H), 3.62 (m, 2H), 3.38 (t, 1H, J=10.5 Hz), 2.75 (m, 2H), 1.44 (s, 9H); $[\alpha]^{29}_{D}$ = +25.8° (c=0.1); HRMS calcd. for $C_{21}H_{25}ClN_2O_4$: 404.1503; found 404.1496.

(S)-Seco-CBI-β-alanine-BOC (**5S**): ^{1}H NMR (300 MHz, $CDCl_3$, 25°C): δ=9.37 (s, 1H), 8.27 (m, 2H), 7.63 (d, 1H, *J*=8.1 Hz), 7.53 (t, 1H, *J*=6.6 Hz), 7.40 (t, 1H, *J*=7.5 Hz), 5.55 (br t, 1H), 4.21 (d, 1H, J=10.2 Hz), 4.06 (t, 1H, J=9.9 Hz), 3.96 (m, 2H), 3.64 (m, 2H),

3.40 (t, *1H*, *J=10.5* Hz), 2.75 (m, 2H), 1.44 (s, 9H); $[\alpha]^{29}_{D}$ = -29.4° (c=0.1); HRMS calcd. for $C_{21}H_{25}ClN_2O_4$: 404.1503; found 404.1490.

***Seco*-CBI-β-alanine-(dimethyl)-γ-amino butyric acid (3R and 3S).** **5** (5 mg, 12.3 μmoles) was deprotected in 3M HCl/ethyl acetate (5 ml) for 30 minutes under argon. The ethyl acetate was removed by evaporation, and the deprotected *seco*-CBI-β-alanine was coevaporated twice from dichloromethane. Separately, a solution of (dimethyl)-γ-aminobutyric acid (4.1 mg, 2 equiv.) in DMF (200 μl) was stirred in a flame dried flask with DCC (25.4 mg, 10 equiv) and N-hydroxysuccinimide (1.4 mg, 1 equiv.) for 1 hour under argon. This was added to the deprotected *seco*-CBI-β-alanine and DIEA (7 μl, 3 equiv.) was added. The solution was stirred for 1 hour under argon, and purified by reversed phase HPLC. *Seco*-CBI-β-alanine-(dimethyl)-g-aminobutyric acid was recovered as a white powder upon lyophilization of the appropriate fraction.

(R)-*Seco*-CBI-β-alanine-(dimethyl)-γ-aminobutyric acid (**3R**): 2.5 mg, 48% (recovery) ^{1}H NMR (300 MHz, CD_3CN, 25°C): δ=10.9 (br s, 1H), 8.12 (d, 1H, *J*=7.5 Hz), 8.02 (s, 1H), 7.69 (d, 1H, *J*=8.4 Hz), 7.48 (t, 1H, *J*=8.4 Hz), 7.33 (t, 1H, *J*=6.9 Hz), 7.09 (br s, 1H), 4.15 (m, 2H), 3.93 (d, 3H), 3.63 (t, 2H), 3.51 (m, 2H), 3.04 (m, 4H), 2.71 (s, 6H), HRMS calcd. for $C_{22}H_{29}ClN_3O_3$ (M+H): 418.1897; found 418.1889.

(S)- *Seco*-CBI-β-alanine-(dimethyl)-γ-aminobutyric acid (**3S**): 3.4 mg, 66% (recovery). ^{1}H NMR (300 MHz, CD_3CN, 25°C): δ=10.9 (br s, 1H), 8.13 (d, 1H, *J*=8.1 Hz), 8.00 (s, 1H), 7.74 (d, 1H, *J*=8.4 Hz), 7.50 (t, 1H, *J*=7.5 Hz), 7.34 (t, 1H, *J*=7.5 Hz), 7.00 (br s, 1H), 4.23 (m, 2H), 4.20 (m, 1H), 3.93 (d, 2H), 3.67 (t, 2H), 3.51 (m, 2H), 3.04 (t, 2H), 2.75 (s, 6H). HRMS calcd. for $C_{22}H_{29}ClN_3O_3$ (M+H): 418.1897; found 418.1907.

DNA Reagents and Materials. Enzymes were purchased from Boehringer-Mannheim and used with their supplied buffers. Deoxyadenosine and thymidine 5'-[α-^{32}P] triphosphates were obtained from Dupont/New England Nuclear, and deoxyadenosine 5'-[γ-^{32}P] triphosphates were purchased from I.C.N. Sonicated, deproteinized calf thymus DNA was acquired from Pharmacia. RNase free water was obtained from USB and used for all reactions. All other reagents and materials were used as received. All DNA manipulations were performed according to standard protocols.

Construction of Plasmid DNA. The plasmid pAC1 was constructed using previously described methods. Fluorescent sequencing was performed at the DNA Sequencing Facility at the California Institute of Technology and was used to verify the presence of the desired insert. Concentration of the prepared plasmid was determined at 260 nm from the relationship of 1 OD unit=50 μg mL^{-1} duplex DNA.

PCR Labeling to generate 5'-End-Labeled Restriction Fragments. Two 21 mer primers were synthesized for PCR amplification: primer A (labeled)5'-AATTCGAGCTCGGTACCCGGG-3' and primer B (unlabeled) 5'-CTGGCACGACAGGTTTCCCGA-3'. Primer A was treated with T4 polynucleotide kinase and deoxyadenosine 5'-[γ-^{32}P] triphosphate as previously described. PCR reactions containing 60 pmol each primer, 10 μl PCR buffer (Boehringer-Mannheim), 3.7 μl template (0.003 μg/mL), 2 μl dNTP mix (each at 10 mM), 1 μl 100X BSA (New England Biolabs) and 83 μl water were heated at 70 °C for 5 minutes. Four units of Taq Polymerase were added (Boehringer-Mannheim). Thirty amplification cycles were performed, each cycle consisting of the following segments: 94 °C for 1 minute, 54 °C for 1 minute, and 72 °C for 1.5 minutes. Following the last cycle, 10 minutes of

extension at 72 °C completed the reaction. The PCR products were gel purified as previously reported for 3'-end labeling protocols.

Cleavage Reactions. All reactions were carried out in a volume of 50 µl. A polyamide or *seco*-CBI-dimethyl gaba stock solution or water (for reference lane) was added to an assay buffer of TE (pH7.5) and 20 kcpm of 3'- or 5'-radiolabeled DNA. The solutions were allowed to equilibrate for 12 hours or the appropriate time (for time course reactions) at 37 °C. The reactions were stopped with 60 µl of a solution containing NaOAc (600 mM), EDTA pH 8.0 (12.5 mM), calf thymus DNA (150 µM base pair), glycogen (0.8 mg mL^{-1}), and NaCl (2 M). Ethanol was added to remove unbound polyamide and precipitate the products. The reactions were resuspended in 20 µl of TE (pH 7.5) and cleavage was initiated by heating at 95 °C for 30 minutes. The cleavage products were precipitated with 150 µl ethanol and then resuspended in 100 mM trisborate-EDTA/80% formamide loading buffer, and denatured and loaded onto polyacrylamide gels as previously reported. The gels were quantitated by the use of storage phosphor technology. Yield or efficiency of alkylation was determined as the ratio between the volume integration assigned to the products and the sum of the volumes of all the products in the lane.

Acknowledgements

We are grateful to the National Institutes of Health for research support and a National Research Service Award to A.Y.C. We thank G.M. Hathaway and the Caltech Protein/Peptide Microanalytical Laboratory for MALDI-TOF mass spectrometry.

References

1. Boger, D. L. & Johnson, D. S. CC-1065 and the duocarmycins: Understanding their biological function through mechanistic studies. *Angew. Chem., Int. Ed. Eng.* **35**, 1438-1474 (1996).

2. Monroe, T. J. & Mitchell, M. A. In-Vivo Mutagenesis Induced by CC-1065 and Adozelesin DNA Alkylation in a Transgenic Mouse Model. *Cancer Res.* **53**, 5690-5696 (1993).

3. McGovren, J. P., Clarke, G. L., Pratt, E. A. & Dekoning, T. F. Preliminary Toxicity Studies with the DNA-Binding Antibiotic, CC-1065. *J. Antibiot.* **37**, 63-70 (1984).

4. Boger, D. L. et al. Reversed and sandwiched analogs of duocarmycin SA: Establishment of the origin of the sequence-selective alkylation of DNA and new insights into the source of catalysis. *J. Am. Chem. Soc.* **119**, 4987-4998 (1997).

5. Boger, D. L., Boyce, C. W. & Johnson, D. S. pH dependence of the rate of DNA alkylation for (+)-duocarmycin SA and (+)-CCBI-TMI. *Bioorg. Med. Chem. Lett.* **7**, 233-238 (1997).

6. Boger, D. L. & Turnbull, P. Synthesis and evaluation of CC-1065 and duocarmycin analogs incorporating the 1,2,3,4,11,11a- hexahydrocyclopropa[c]naphtho [2,1-b]azepin-6-one (CNA) alkylation subunit: Structural features that govern reactivity and reaction regioselectivity. *J. Org. Chem.* **62**, 5849-5863 (1997).

7. Smith, J. A. et al. The structural basis for in situ activation of DNA alkylation by duocarmycin SA. *J. Mol. Biol.* **300**, 1195-1204 (2000).

8. Eis, P. S. et al. High resolution solution structure of a DNA duplex alkylated by the antitumor agent duocarmycin SA. *J. Mol. Biol.* **272**, 237-252 (1997).

9. Boger, D. L. & Garbaccio, R. M. Shape-dependent catalysis: Insights into the source of catalysis for the CC-1065 and duocarmycin DNA alkylation reaction. *Acc. Chem. Res.* **32**, 1043-1052 (1999).

10. Boger, D. L., Yun, W. Y. & Han, N. H. 1,2,9,9a-Tetrahydrocyclopropa[c]Benz [e]Indol-4-One (CBI) Analogs of CC-1065 and the Duocarmycins - Synthesis and Evaluation. *Bioorg. Med. Chem.* **3**, 1429-1453 (1995).

11. Boger, D. L., Ishizaki, T., Zarrinmayeh, H., Kitos, P. A. & Suntornwat, O. A Potent, Simple Derivative of an Analog of the CC-1065 Alkylation Subunit. *Bioorg. Med. Chem. Lett.* **1**, 55-58 (1991).

12. White, S., Szewczyk, J. W., Turner, J. M., Baird, E. E. & Dervan, P. B. Recognition of the four Watson-Crick base pairs in the DNA minor groove by synthetic ligands. *Nature* **391**, 468-471 (1998).

13. Herman, D. M., Baird, E. E. & Dervan, P. B. Stereochemical control of the DNA binding affinity, sequence specificity, and orientation preference of chiral hairpin polyamides in the minor groove. *J. Am. Chem. Soc.* **120**, 1382-1391 (1998).

14. Boger, D. L., Ishizaki, T., Kitos, P. A. & Suntornwat, O. Synthesis of N-(Tert-Butyloxycarbonyl)-CBI, CBI, CBI-$CDPI_1$, and CBI-$CDPI_2$ - Enhanced Functional Analogs of Cc-1065 Incorporating the 1,2,9,9a-Tetrahydrocyclopropa [c]Benz [e]Indol- 4-One (CBI) Left-Hand Subunit. *J. Org. Chem.* **55**, 5823-5832 (1990).

15. *Boger, D. L., Yun, W. Y. & Teegarden, B. R. An Improved Synthesis of 1,2,9,9a-Tetrahydrocyclopropa[c]Benz[e]Indol-4-One (CBI) - a Simplified Analog of the CC-1065 Alkylation Subunit. *J. Org. Chem.* **57**, 2873-2876 (1992).

16. Boger, D. L. & McKie, J. A. An Efficient Synthesis of 1,2,9,9a- Tetrahydrocyclopropa[c]Benz[e]Indol-4-One (CBI) - an Enhanced and Simplified Analog of the CC-1065 and Duocarmycin Alkylation Subunits. *J. Org. Chem.* **60**, 1271-1275 (1995).

17. Boger, D. L. & Yun, W. Y. Cbi-Tmi - Synthesis and Evaluation of a Key Analog of the Duocarmycins - Validation of a Direct Relationship between Chemical Solvolytic Stability and Cytotoxic Potency and Confirmation of the Structural Features Responsible for the Distinguishing Behavior of Enantiomeric Pairs of Agents. *J. Am. Chem. Soc.* **116**, 7996-8006 (1994).

18. Baird, E. E. & Dervan, P. B. Solid phase synthesis of polyamides containing imidazole and pyrrole amino acids. *J. Am. Chem. Soc.* **118**, 6141-6146 (1996).

19. Boger, D. L. & Munk, S. A. Dna Alkylation Properties of Enhanced Functional Analogs of CC- 1065 Incorporating the 1,2,9,9a-Tetrahydrocyclopropa[1,2-c]Benz[1,2-e]Indol-4-One (CBI) Alkylation Subunit. *J. Am. Chem. Soc.* **114**, 5487-5496 (1992).

20. Boger, D. L. et al. DNA Alkylation Properties of the Duocarmycins - (+)-Duocarmycin-A, Epi-(+)-Duocarmycin-A, Ent-(-)-Duocarmycin-A and Epi,Ent-(-)-Duocarmycin-A. *Bioorg. Med. Chem. Lett.* **2**, 759-& (1992).

21. *Lukhtanov*, E. A. et al. Sequence and structure dependence of the hybridization-triggered reaction of oligonucleotides bearing conjugated cyclopropapyrroloindole. *J. Am. Chem. Soc.* **119**, 6214-6225 (1997).

22. Sugiyama, H., Lian, C. Y., Isomura, M., Saito, I. & Wang, A. H. J. Distamycin A modulates the sequence specificity of DNA alkylation by duocarmycin A. *Proc. Natl. Acad. Sci. USA* **93**, 14405-14410 (1996).

23. Tao, Z. F., Fujiwara, T., Saito, I. & Sugiyama, H. Rational design of sequence-specific DNA alkylating agents based on duocarmycin A and pyrrole-imidazole hairpin polyamides. *J. Am. Chem. Soc.* **121**, 4961-4967 (1999).

24. Ebbinghaus, S. W., Fortinberry, H. & Gamper, H. B. Inhibition of transcription elongation in the HER-2/neu coding sequence by triplex-directed covalent modification of the template strand. *Biochemistry* **38**, 619-628 (1999).

CHAPTER THREE

Utilization of Polyamide-*seco*-CBI Conjugates to Stimulate DNA Repair

Abstract

Pyrrole-Imidazole (Py-Im) polyamides are cell permeable ligands that can be targeted to any predetermined DNA sequence. Py-Im polyamide-*seco*-CBI conjugates are capable of cleaving DNA in high yields at targeted adenines proximal to the polyamide binding site. Because of the generality of the pairing rules, and polyamides demonstrated utility in *in vivo* systems, these conjugates may be useful for studying the effects of this class of DNA binding ligands on homologous recombination and DNA repair. Triplex oligonucleotides have previously been shown to influence homologous recombination and directed mutagenesis. The effects of other DNA binding ligands on these processes have not been established. Described are the design and synthesis of Py-Im polyamide-*seco*-CBI conjugates that cleave DNA at a determined site on a specific gene. Currently, these compounds are being studied *in vivo* in hopes that DNA cleavage will stimulate a DNA repair event and homologous recombination. If successful, these studies could have future applications in the areas of gene therapy.

Background

Using DNA-binding Ligands to Study DNA Repair Processes

Synthetic ligands that can target specific sequences of DNA and bind with high affinity are currently studied to manipulate transcription in both *in vitro* and *in vivo* systems. Previously were described efforts to inhibit or activate transcription by using Py-Im polyamides to compete with the binding of required transcription factors.[1,2] Other endeavors in this regard involve using such ligands to interfere with transcription elongation through covalent attachment to the DNA via a reactive moiety linked to the DNA-binding ligand. This approach was discussed concerning Py-Im polyamides conjugates with *seco*-CBI.[3]

Another use for the study of molecules that target specific sequences of DNA is in the area of gene repair. Researchers often use DNA damaging agents to study how cells repair such damage, and how efficient and accurate these repair processes are. One problem with the majority of DNA damaging agents is that they have broad sequence selectivities, making it difficult to study the effects on a specific gene. There is much interest in being able to target a specific gene and study how the cell can repair such damage, and if that repair is accurate.[4] While the majority of the time accurate repair of DNA damage is desirable, there may be instances where an error in DNA repair can yield a beneficial result. This idea of directed mutagenesis is one approach to gene therapy. If it were possible that a gene specific ligand could induce a favorable mutation during the repair process, then the genome of that cell would be permanently altered for the better. Understanding the processes that guide and direct DNA repair will ultimately have implications in this discipline.

Much of the research in using DNA-binding ligands to study DNA repair processes has used triplex oligonucleotides.[5,6] There are well established rules of triplex formation and their recognition sequences, and they are easily synthesized and accessible. However, triplex oligonucleotides face problems in delivery because they are anionic polymers, and in stability because they are subject to degradation by nucleases. The recognition sequences are often restricted to purine tracts which limit their use in any given gene. The merits of using Py-Im polyamides to study DNA repair are evident in the generality of the pairing rules[7], their cell permeability[1,8], and accessibility through solid phase synthesis.[9]

To study DNA repair and recombination, we proposed using Py-Im polyamides to direct known DNA damaging agents to specific genes. The lack of sequence specificity in traditional DNA alkylating agents such as nitrogen mustards, CC-1065, and mitomycin make them unsuitable for a gene specific study of DNA repair. The use of polyamide conjugates allows us to control the damage inflicted in order to stimulate repair and recombination to the particular gene in study. In collaboration with Dr. Matt Porteus of the Baltimore group, experiments were designed to use polyamide conjugates of *seco*-CBI, nitrogen mustards (Mr. Nicholas Wurtz of the Dervan group), and camptothecin (Mr. Clay Wang of the Dervan group) to stimulate DNA repair and homologous recombination. Described here are the efforts in using *seco*-CBI polyamide conjugates for these experiments.

Experimental Design

The experimental design was based on observations in the Baltimore group using Green Fluorescent Protein (GFP). When a mutant GFP gene with an internal stop codon

is expressed, functional GFP is not detected (Figure 3.1a). When substrate DNA is added containing the wild type gene is added, homologous recombination occurs such that functional GFP can be detected for 1 in every 10^6 cells (Figure 3.1b). If instead of the internal stop codon, a site for a rare-cutting restriction enzyme is inserted[4], a double strand break can be generated. For the case of homologous recombination with the substrate for the double strand break, functional GFP can be detected for 1 in every 500 cells (Figure 3.1c). Thus these experiments were designed to address the question of whether or not polyamides could be used to incorporate a double strand break, and if so could we generate similar results of recombination and mutagenesis (Figure 3.1d).

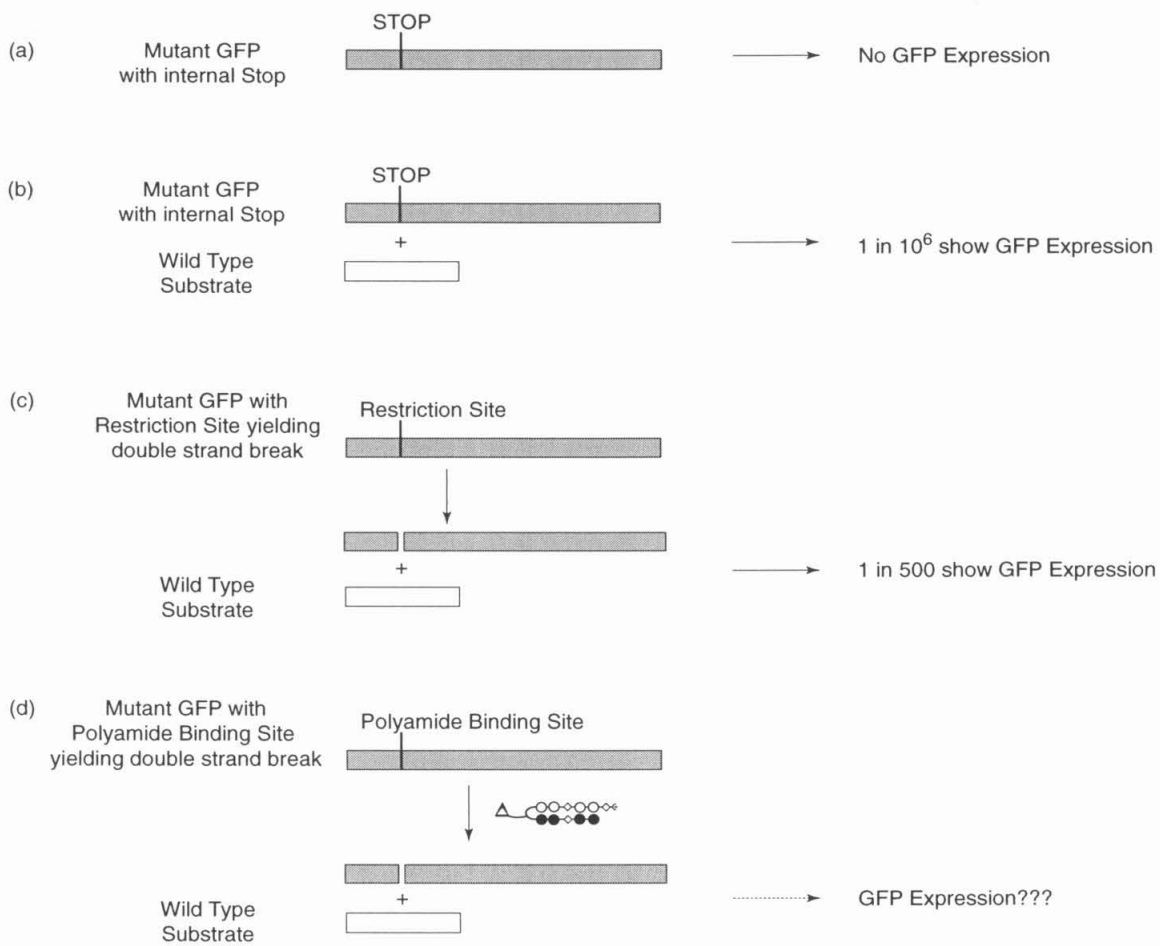


Figure 3.1. Experimental design for proposed polyamide conjugates. See text for further details.

Nick Wurtz of the Dervan group chose three polyamides to examine in these experiments (Figure 3.2). Two constructs were designed to incorporate these polyamides. The first uses the two hairpin dimers ImPy-β-ImPy-$\gamma^{(R)\text{-CBI}}$-Py-β-ImImPy-β-Dp (**1R**) and ImIm-β-ImIm-γ-$^{(R)\text{-CBI}}$-PyPy-β-PyPy-β-Dp (**2R**). The second uses the homodimer ImImPyPy-β-PyPyPyPy-β-TA-(S)-CBI (**3S**). The polyamides selected have have been previously characterized and have high binding affinities and good specificities to mismatch.[1,10] The enantiomer of CBI chosen was determined by the insert design. Earlier studies showed that the strand selectivity of polyamide-*seco*-CBI conjugates is dependent on the enantiomer chosen. For the dimer insert, polyamides **1** and **2** are designed in a 5' to 3' orientation from the alkylation site. Thus, for these polyamides, the (R)-enantiomer of CBI was chosen. This orientation is consistent with that of the parent alkylating agent and as well as with the previous studies of polyamide-*seco*-CBI conjugates. For the homodimer insert, each half of the homodimer needs to bind in a 3' to 5' orientation from the alkylation site, thus dictating the need for the (S)-enantiomer of *seco*-CBI.

The insert containing the binding sites of the polyamides for each construct (dimer and homodimer) was inserted in the middle of the GFP gene (Figure 3.3). The insert includes a stop codon so that cells that do not undergo recombination will not express functional GFP. In the event of a double strand break, repair processes could correct the mutation, which would then be detectable by fluorescence. If the polyamides alkylate as designed to do, they will in theory act the same way as a rare-cutting endonuclease would, causing a double strand break. It is unclear what the effect of the overhang will have on the repair mechanism of the alkylated adduct.[4] Rare-cutting

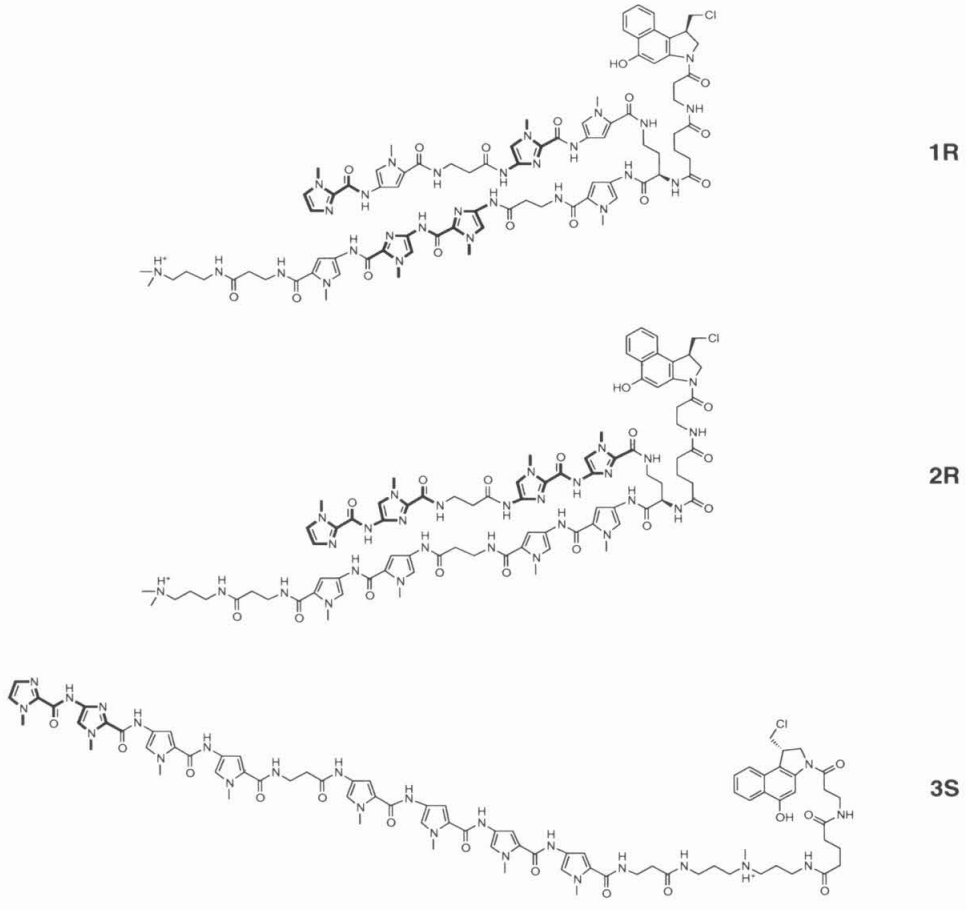


Figure 3.2. Structures of the polyamide conjugates used for this study. ImPy-β-ImPy-γ$^{(R)\text{-}CBI}$-Py-β-ImImPy-β-Dp (**1R**), ImIm-β-ImIm-γ-$^{(R)\text{-}CBI}$-PyPy-β-PyPy-β-Dp (**2R**), and ImImPyPy-β-PyPyPyPy-β-TA-(S)-CBI (**3S**).

endoncleases typically either leave a blunt end or a 2 to 4 base overhang to either the 3' or 5' side. In the case of the dimers, the designed overhang will be similar to that of an endonuclease. In the case of the homodimers, there is a fairly substantial gap between the two cleavage sites of duplex DNA (19 bases). If there are differences observed between the two systems, a next generation design might try to incorporate a range of overhangs to see if a trend can be established.

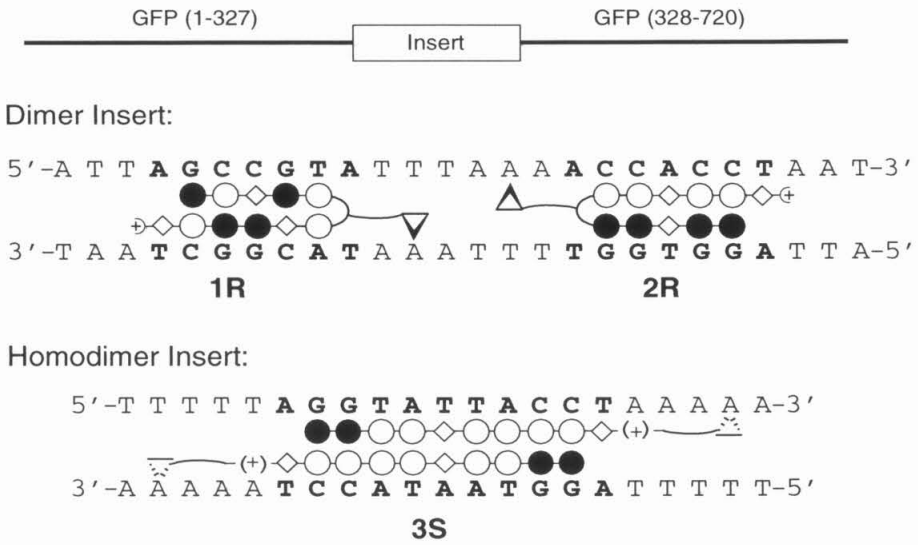


Figure 3.3. DNA construct of dimer and homodimer inserts for GFP experiments.

Results

Synthesis of Polyamide seco-CBI-Conjugates

The parent unmodified polyamides **1**, **2**, and **3** were synthesized on solid support as previously reported.[9] Polyamides **1** and **2** were synthesized with (R)-2,4-diaminobutyric acid (DABA) in the turn position.[11] A succinimidyl ester was attached to the amine of the turn to yield **1-DSG**. Subsequent addition of (R)-β-ala-*seco*-CBI yielded the desired polyamide-*seco*-CBI conjugate.[3] A representative conjugate synthesis is shown in Figure 3.4. Polyamide **3** was cleaved from the resin with 3,3'-diamino-N-methyldipropylamine (triamine, TA) and the conjugates were then synthesized as polyamides **1R** and **2R** with the (S)-enantiomer of *seco*-CBI.[12]

Thermally Induced Cleavage Reactions

Polyamide **1R** was analyzed on a 221 base pair fragment of GFP containing the designed insert (Figure 3.5a). Analysis of the cleavage reactions reveal that polyamide **1R** alkylates strongly at the adenine two base pairs removed from the designed match

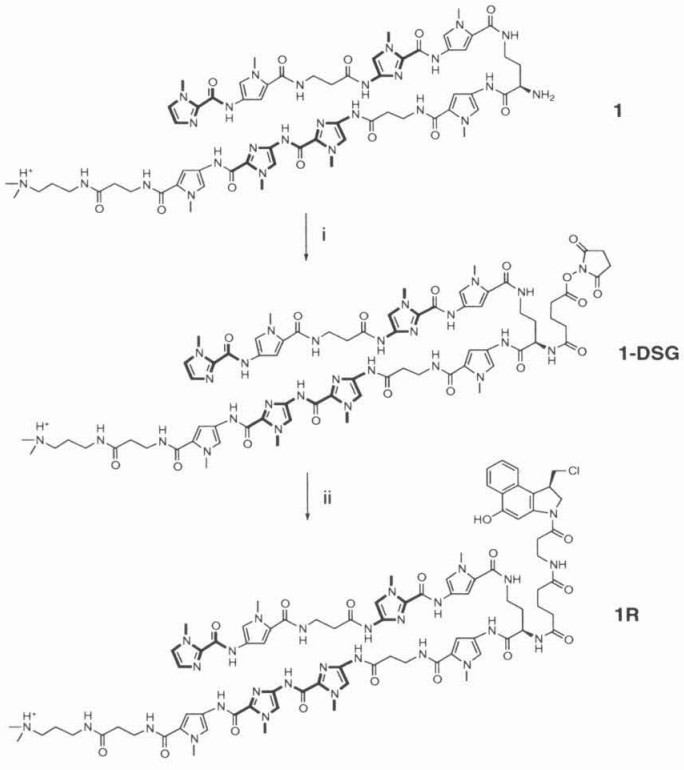


Figure 3.4. Synthesis of Polyamide-*seco*-CBI conjugates. (i) DSG (10 equiv.), (ii) (R)-*seco*-CBI-β-ala.

a) GFP Dimer Fragment (177 bp+ 44 bp-insert)

```
5'-AAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTC
3'-TTCGTCGTGCTGAAGAAGTTCAGGCGGTACGGGCTTCCGATGCAGGTCCTCGCGTGGTAGAAGAAG

AAGGACGACGGCAACTACAAGACC-TAAGCTCTCGAGATTAGCCGTATTTAAAACCACCTAATAAGCTT-
TTCCTGCTGCCGTTGATGTTCTGG-ATTCGAGAGCTCTAATCGGCATAAATTTTGGTGGATTATTCGAA-

CGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAG
GCGCGGCTCCACTTCAAGCTCCCGCTGTGGGACCACTTGGCGTAGCTCGACTTCCCGTAGCTGAAGTTC

GAGGACGGCAACATCCTG-3'
CTCCTGCCGTTGTAGGAC-5'
```

b) GFP Homodimer Fragment (177 bp+ 38 bp-insert)

```
5'-AAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTC
3'-TTCGTCGTGCTGAAGAAGTTCAGGCGGTACGGGCTTCCGATGCAGGTCCTCGCGTGGTAGAAGAAG

AAGGACGACGGCAACTACAAGACC-TAAGCTCTCGAGTTTTTAGGTATTACCTAAAAAAGCTT-CGCGC
TTCCTGCTGCCGTTGATGTTCTGG-ATTCGAGAGCTCAAAAATCCATAATGGATTTTTTTCGAA-GCGCG

CGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGA
GCTCCACTTCAAGCTCCCGCTGTGGGACCACTTGGCGTAGCTCGACTTCCCGTAGCTGAAGTTCCTCCT

CGGCAACATCCTG-3'
GCCGTTGTAGGAC-5'
```

Figure 3.5. GFP fragments used in this study. Insert with polyamide binding sites are between the dashes, polyamide binding sites are in bold. (a) dimer construct, (b) homodimer construct.

site, 5'-*AGCCGTA*-3' on the bottom strand (Figure 3.6b). At 1 nM concentrations, **1R** cleaves this match site with a yield of 79%. On this strand, excellent specificity is shown, as 2 mismatch alkylation sites only begin to appear at 10 nM concentrations. These sites correspond to a double base pair mismatch 5'-CGCCGAG-3' (mismatch underlined) and a second double base pair mismatch 5'-AGGTGAA-3'which have cleavage yields of 13% and 9% respectively at 10 nM concentrations. The opposite strand (top strand) of this fragment was also examined to determine if there were mismatch alkylation sites on this strand as well (Figure 3.6a). On this strand, polyamide **1R** shows one strong alkylation site at a single base pair mismatch site 5'-TGCCGTC-3'. The yield of alkylation at this site is 79% at 3 nM concentrations of polyamide **1R**.

Polyamide **2R** was examined on the same 221 base pair fragment as polyamide **1R** (Figure 3.5a). On the top strand proximal to the intended match site 5'-AGGTGGT-3', polyamide **2R** shows a strong alkylation site which starts to appear at 30 pM (Figure 3.7a). At 1 nM concentrations polyamide **2R**, the cleavage yield is 86% at this site. On the same strand at 3 nM concentrations, alkylation is observed proximal to a double base pair mismatch site 5'-AGGTCTT-3'. The cleavage yield at this site is 19% at 3 nM and 56% at 10 nM concentrations. On the bottom strand, one alkylation site is seen at a single base pair mismatch site 5'-AGGTGAA-3'(Figure 3.7b). The alkylation band appears at 3 nM concentrations and cleaves with 75% yield at 10 nM concentrations.

Polyamide **3S** was assayed on a 215 base pair fragment containing the homodimer polyamide match site (Figure 3.5b). Both strands show alkylation at the designed match site 5'-AGGTATTACCT-3'. For both strands, alkylation is seen starting at 300 pM (Figure 3.8). The top strand alkylation seems more specific, which has a cleavage yield

Figure 3.6. Thermally induced strand cleavage on the 5'-end labeled and 3'-end labeled 221 base pair restriction fragment by ImPy-β-ImPy-$\gamma^{(R\text{-}seco\text{-}CBI)}$-Py-β-ImImPy-β-Dp (**1R**). Storage phosphor autoradiograms of 8% denaturing polyacrylamide gels used to separate the fragments generated by heat induced DNA cleavage at alkylation sites. All lanes contain 10 kcpm of either 5' radiolabeled DNA. Each reaction was equilibrated in TE, pH 7.5 at 37 °C for 12 H. The unbound polyamide was removed by precipitation, and then strand cleavage was induced by heating at 95 °C for 30 min. (a) 5'-^{32}P-end labeled restriction fragment-top strand. (b) 5'-^{32}P-end labeled restriction fragment-bottom strand. (a-b) lane 1, intact DNA; lanes 2-8, 10 pM, 30 pM, 100 pM, 300 pM, 1 nM, 3 nM, 10 nM respectively of the corresponding polyamide; lane 9, A-specific reaction; lane 10, G-specific reaction. (a-b) Match site 5'-AGCCGTA-3' and single base pair mismatch site 5'-GACGGCA-3' are indicated in bold on the sequence, with arrows indicating cleavage bands.

1R: ImPy-β-ImPy-$\gamma^{(R)\text{-CBI}}$-Py-β-ImImPy-β-Dp

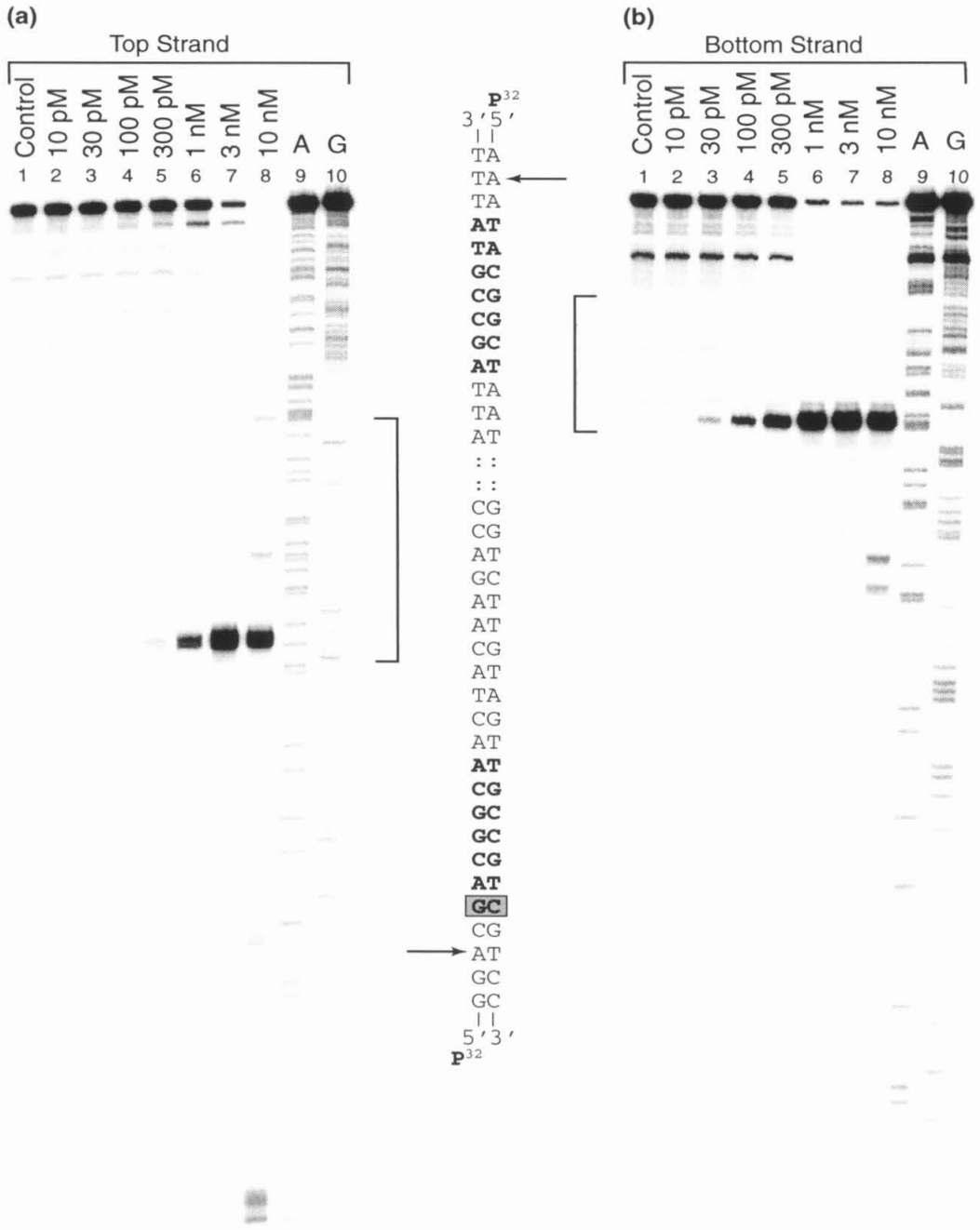


Figure 3.7 Thermally induced strand cleavage on the 5'-end labeled and 3'-end labeled 221 base pair restriction fragment by ImIm-β-ImIm-$\gamma^{(R\text{-}seco\text{-}CBI)}$-PyPy-β-PyPy-β-Dp (**2R**). Storage phosphor autoradiograms of 8% denaturing polyacrylamide gels used to separate the fragments generated by heat induced DNA cleavage at alkylation sites. All lanes contain 10 kcpm of either 5' radiolabeled DNA. Each reaction was equilibrated in TE, pH 7.5 at 37 °C for 12 H. The unbound polyamide was removed by precipitation, and then strand cleavage was induced by heating at 95 °C for 30 min. (a) 5'-^{32}P-end labeled restriction fragment-top strand. (b) 5'-^{32}P-end labeled restriction fragment-bottom strand. (a-b) lane 1, intact DNA; lanes 2-8, 10 pM, 30 pM, 100 pM, 300 pM, 1 nM, 3 nM, 10 nM respectively of the corresponding polyamide; lane 9, A-specific reaction; lane 10, G-specific reaction. (a-b) Match site 5'-AGGTGGT-3' and single base pair mismatch site 5'-AGGTCTT-3' and 5'-AGGTGAA-3' are indicated in bold on the sequence, with arrows indicating cleavage bands.

2R: ImIm-β-ImIm-$\gamma^{(R)\text{-}CBI}$-PyPy-β-PyPy-β-Dp

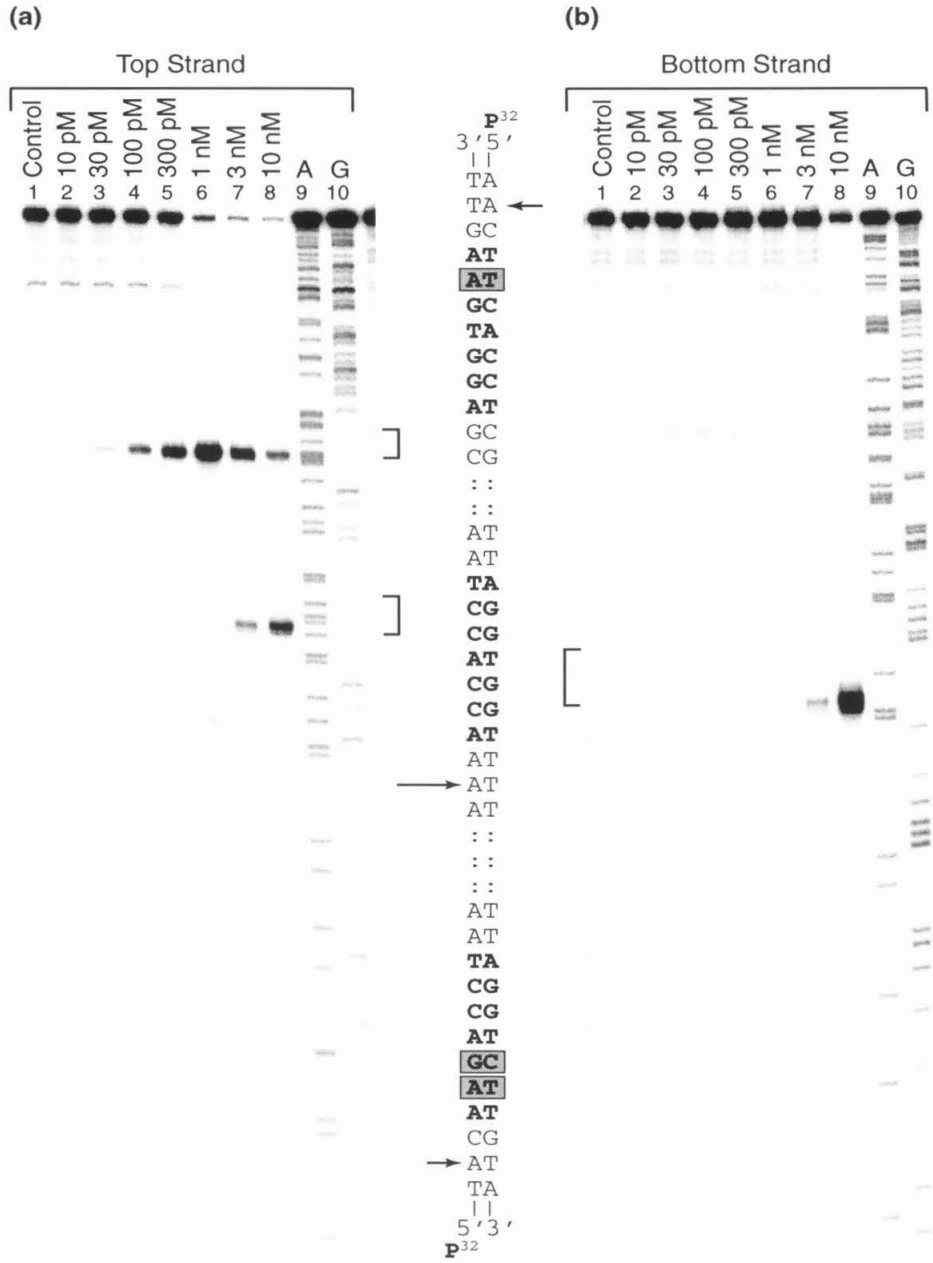
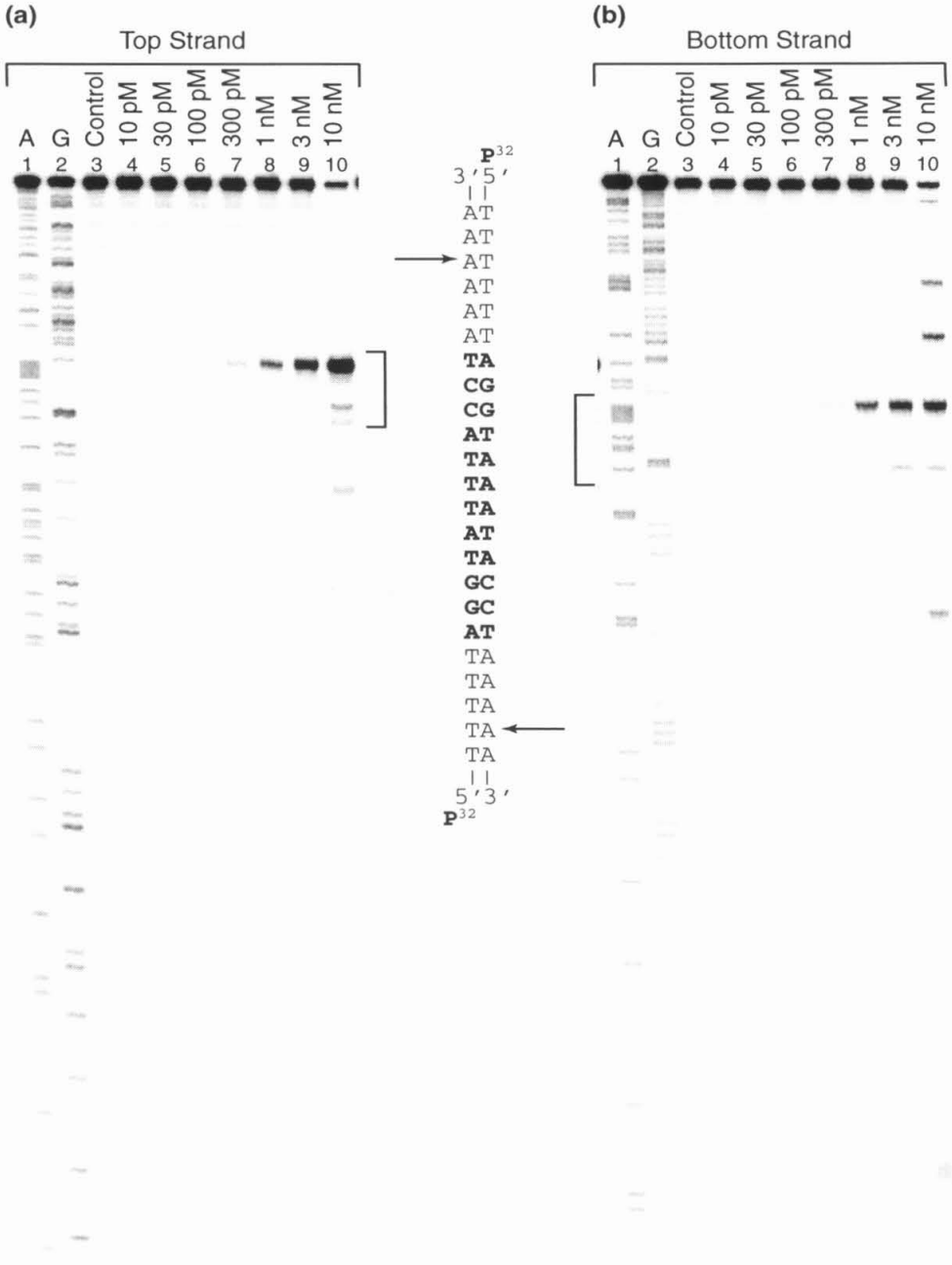


Figure 3.8 Thermally induced strand cleavage on the 5'-end labeled and 3'-end labeled 215 base pair restriction fragment by ImImPyPy-β-PyPyPyPy-β-TA-(S)-CBI (**3S**). Storage phosphor autoradiograms of 8% denaturing polyacrylamide gels used to separate the fragments generated by heat induced DNA cleavage at alkylation sites. All lanes contain 10 kcpm of either 5' radiolabeled DNA. Each reaction was equilibrated in TE, pH 7.5 at 37 °C for 12 H. The unbound polyamide was removed by precipitation, and then strand cleavage was induced by heating at 95 °C for 30 min. (a) 5'-^{32}P-end labeled restriction fragment-top strand. (b) 5'-^{32}P-end labeled restriction fragment-bottom strand. (a-b) lane 1, A-specific reaction; lane 2, G-specific reaction; lane 3, intact DNA lanes 4-10, 10 pM, 30 pM, 100 pM, 300 pM, 1 nM, 3 nM, 10 nM respectively of the corresponding polyamide. (a-b) Match site 5'-AGGTATTTACCT-3' is indicated in bold on the sequence, with arrows indicating cleavage bands.

3S: ImImPyPy-β-PyPyPyPy-β-TA-(S)-CBI



of 69% at the target site at 10 nM concentrations of polyamide **3S** (Figure 3.8a). There are other alkylation bands on this strand, none of which account for more than 5% of the total DNA in the lane. Alkylation on the bottom strand was seen at a lower yield, 30% yield at the match site at 10 nM concentrations of polyamide **3S** (Figure 3.8b). Other mismatch alkylation sites show yields as high as 10%-15% at the same concentrations. Polyamide **3S** was not as soluble as the hairpin polyamides **1R** and **2R**, which may account for the higher concentrations needed to observe alkylation.

A summary of the alkylation sites for the dimer and homodimer sites are illustrated in Figure 3.9. At 1 nM concentrations of each of the hairpin polyamides **1R** and **2R**, the majority of cleavage sites are in the designed region. The use of these two polyamides together should afford a double strand break with a three base pair overhang. There is also a large contribution from single base pair mismatch alkylation site for **1R**. It is unclear how this alkylation site will affect the *in vivo* assays. The homodimer **3S** alkylates fairly specifically at 3 nM concentrations, but at lower yields than the hairpin polyamide conjugates. It will be interesting to compare the results of these two systems to determine the effects, if any, of the *in vitro* yields versus *in vivo* result, issues of solubility and permeability, and the nature of the double strand break (a few base overhang versus >15 bases).

Gene Correction Experiments

Currently, Dr. Matt Porteus of the Baltimore group is working to set up these experiments to study the effects of Py-Im polyamides on homologous recombination. It was hoped that upon treatment with polyamide-alkylator conjugates, cells that had undergone recombination or repair would be detected by expression of GFP. The results

of these experiments thus far have shown that the polyamide-*seco*-CBI conjugates were toxic to cells at 500 nM and 1 μM concentrations. When the levels of RNA production are measured, the global levels of RNA are reduced and no gene specific reduction in RNA synthesis is observed.

In order to further study these phenomena, Dr. Porteus has designed a second assay to examine the repair of the DNA damage inflicted by polyamide-alkylator conjugates. His second assay uses a β-galactosidase (lacZ) system (Figure 3.10). The lacZ gene with an internal frameshift is not catalytically active and thus will not show color change upon treatment with X-gal and IPTG. Just by the cellular repair machinery alone (without added wild type substrate and recombination events), 1 in every 5000 cells will change color, indicating that repair has restored the activity of lacZ. If the polyamide-conjugate binding site is placed at the site of the frameshift, upon repair of the lesion formed by the polyamide-*seco*-CBI conjugate, a certain number of the cells should show restored lacZ activity. The first experiments will use plasmid DNA incubated with polyamide conjugate. The plasmid DNA is then transfected into cells and the cell's own repair machinery will repair the lesions, restoring lacZ activity. If these plasmid DNA experiments are successful, Dr. Porteus will incorporate the lacZ gene with polyamide binding sites into the genomic DNA of the cells, and repeat the previous plasmid experiments. The increased sensitivity of this assay to repair will be useful for determining the mechanisms used by the cell to repair damage inflicted by polyamide-alkylator conjugates.

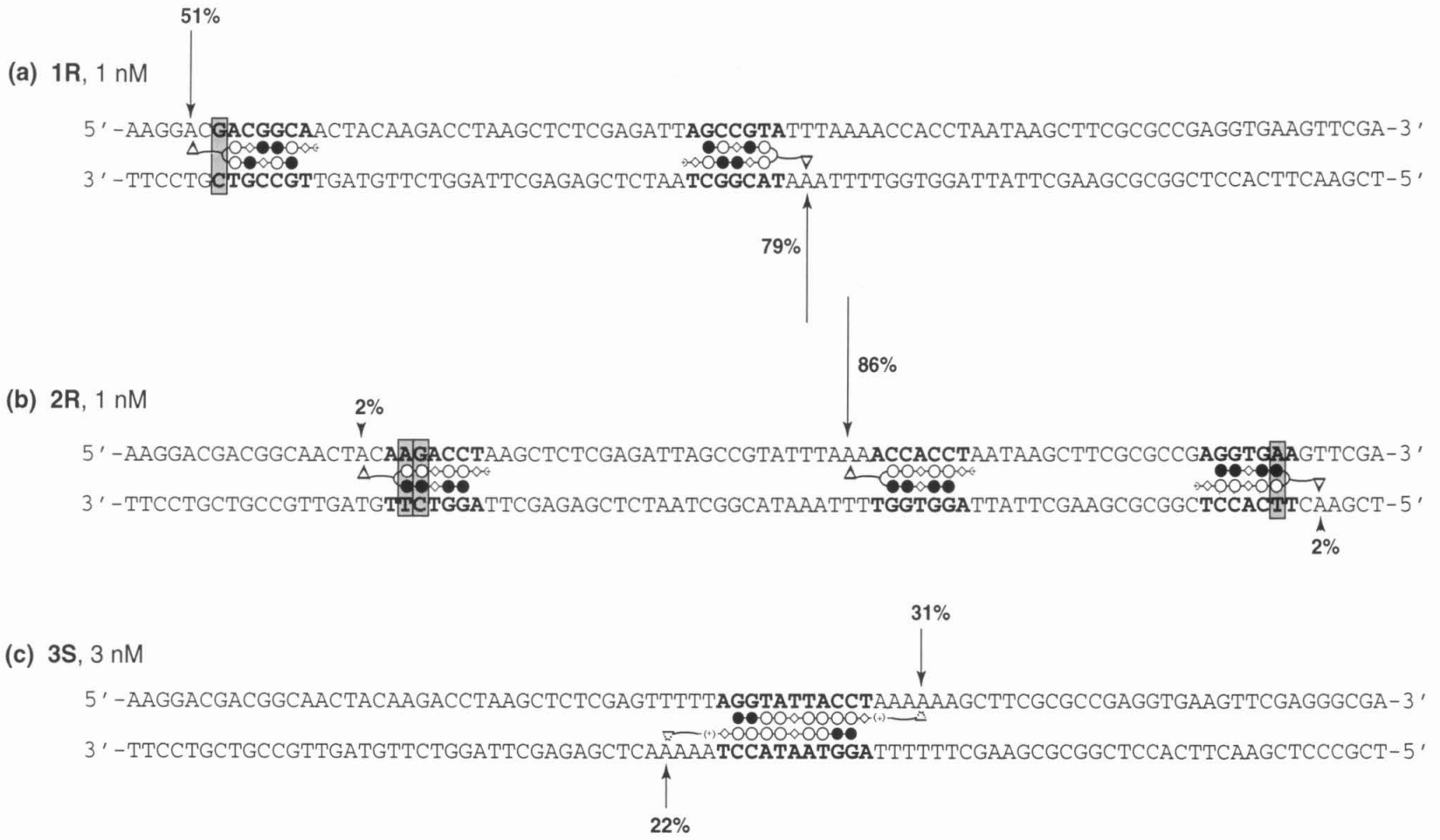


Figure 3.9. Alkylation map indicating alkylation patterns for (a) **1R**, (b) **2R**, and (c) **3S**.

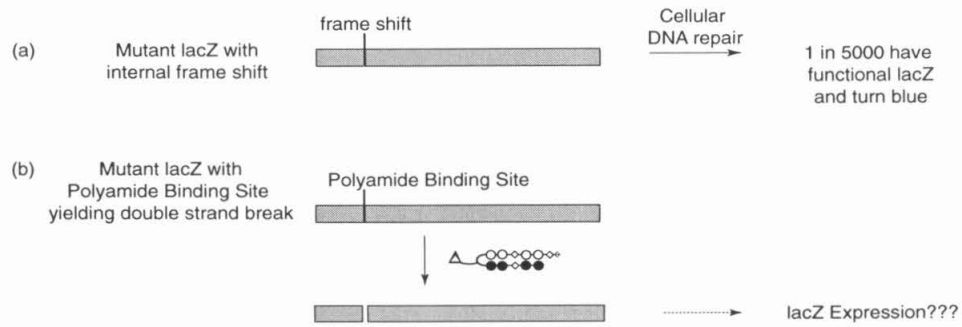


Figure 3.10. LacZ experiments for detecting DNA repair by polyamide-*seco*-CBI conjugate damaged DNA. See text for further details.

Conclusions

The design of Py-Im polyamide conjugates with *seco*-CBI incorporates the ability to target specific DNA sequences and cause a covalent reaction on the DNA in a sequence specific fashion. We believe that this class of molecules, in concert with other polyamide conjugates with nitrogen mustards[13], camptothecin, mitomycin, and PNAs, will provide powerful tools for studying and manipulating transcription in the cell. The results of the forthcoming *in vivo* experiments, whether positive or negative, will shape the design of additional motifs for these classes of compounds and point to new directions to apply these molecules to.

Experimental

Materials

^{1}H NMR spectra were recorded on a General Electric-QE NMR spectrometer at 300 MHz and a Varian Inova NMR spectrometer at 500 MHz with chemical shifts reported in parts per million relative to residual solvent. UV spectra were measured in water on a Hewlett-Packard Model 8452A diode array spectrophotometer. Matrix-assisted, laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF) was performed at

the Protein and Peptide Microanalytical Facility at the California Institute of Technology. Preparatory reversed phase HPLC was performed on a Beckman HPLC with a Waters DeltaPak 25 × 100 mm, 300 Å C18 column equipped with a guard, 0.1% (wt/v) TFA, 0.25% acetonitrile/min.

Synthesis of CBI-polyamide conjugates

ImPy-β-ImPy-(R)^{H_2N}γ-Py-β-ImImPy-β-Dp (1). ImPy-β-ImPy-(R)^{H_2N}γ-Py-β-ImImPy-β-Pam resin was synthesized in a stepwise fashion by Boc-chemistry manual solid phase protocols. A sample of resin was treated with neat (dimethylamino)-propylamine (2 ml), heated (55 °C, 24 hours) and purified by reversed phase HPLC. ImPy-β-ImPy-(R)^{H_2N}γ-Py-β-ImImPy-β-Dp was recovered as a white powder upon lyophilization of the appropriate fraction (14.4 mg, 10.4 μmoles, 6.8% recovery). UV (H_2O) λ_{max} (ε), 312 nm, (66, 600); MALDI-TOF-MS (monoisotopic) [M+H] 1381.70 (calculated 1380.64 for $C_{62}H_{80}N_{26}O_{12}$).

ImPy-β-ImPy-(R)$^{Glu\text{-}NHS}$γ-Py-β-ImImPy-β-Dp (1-DSG). To a solution of disuccinimidyl glutarate (41.9 mg, 120 μmoles) in 2.5 ml DMF was added 100 μl of a 14.3 mM solution of **1** (15.9 mg, 12.8 μmoles) in DMF (800 μl) and DIEA (100 μl). 100 μ l of the solution was added every 15 minutes while stirring. Following the completion of the addition of **1**, the reaction was stirred for 2 hours. The reaction was diluted with 0.1% TFA (15 ml) and the reaction was purified by Sep Pak C_{18}. ImPy-β-ImPy-(R)$^{Glu\text{-}NHS}$γ-Py-β-ImImPy-β-Dp was recovered as a white powder upon lyophilization of the appropriate fraction. UV (H_2O) λ_{max} (ε), 312nm, (66, 600).

ImPy-β-ImPy-(R)$^{(R)\text{-}CBI}$γ-Py-β-ImImPy-β-Dp (1R). To a solution of **1-DSG** (1.8 mg, 1.12 μmoles) in dry DMA was added 30 μl of a 1μM solution of DCC (10 equiv.) in

DMA and 6 µl of a 0.5 M solution of N-hydroxysuccinimide (1 equiv.) in DMA. The solution was stirred for 2 hours. Separately, a solution of *seco*-CBI-Boc-β-ala (0.45 mg, 1 equiv.) was deprotected with 3M HCl/ethyl acetate (5 ml) for 30 minutes under argon. The ethyl acetate was then removed by vacuum and then coevaporated twice from dichloromethane to yield *seco*-CBI-β-ala amine. The gray solid was then dissolved in 50 µl of dry DMF and then added to the polyamide solution. DIEA (2 µl, 10 equiv.) was then added and the reaction was stirred for 3 hours under argon. Upon completion, the reaction was diluted with 0.1% TFA (2 ml) and the reaction was purified by reversed phase HPLC. ImPy-β-ImPy-(R)$^{(R)\text{-CBI}}$γ-Py-β-ImImPy-β-Dp was recovered as a white powder upon lyophilization of the appropriate fraction (0.6 mg, 337 nM, 30% recovery). UV (H_2O) λ_{max} (ε), 314 nm, (73, 854); MALDI-TOF-MS (monoisotopic) [M+H] 1781.9 (calculated 1780.76 for $C_{83}H_{101}ClN_{28}O_{16}$).

ImIm-β-ImIm-(R)^{H_2N}γ-PyPy-β-PyPy-β-Dp (2). ImIm-β-ImIm-(R)^{H_2N}γ-PyPy-β-PyPy-β-Pam resin was synthesized in a stepwise fashion by Boc-chemistry manual solid phase protocols. A sample of resin was treated with neat (dimethylamino)-propylamine (2 ml), heated (55 °C, 24 hours) and purified by reversed phase HPLC. ImIm-β-ImIm-(R)^{H_2N}γ-PyPy-β-PyPy-β-Dp was recovered as a white powder upon lyophilization of the appropriate fraction (12.9 mg, 9.3 µmoles, 12.1% recovery). UV (H_2O) λ_{max} (ε), 312 nm, (66, 600); MALDI-TOF-MS (monoisotopic) [M+H] 1381.76 (calculated 1380.64 for $C_{62}H_{80}N_{26}O_{12}$).

ImIm-β-ImIm-(R)$^{Glu\text{-}NHS}$γ-PyPy-β-PyPy-β-Dp (2-DSG). 2-DSG was prepared in an analogous fashion to **1-DSG** from **2**. ImIm-β-ImIm-(R)$^{Glu\text{-}NHS}$γ-PyPy-β-PyPy-β-Dp was recovered as a white powder upon lyophilization of the appropriate fraction. UV (H_2O)

λ_{max} (ε), *312nm*, (66, 600).

ImIm-β-ImIm-(R)$^{(R)\text{-}CBI}$γ-PyPy-β-PyPy-β-Dp (2R). 2R was prepared in an analogous fashion to **1R** from **2-DSG**. ImIm-β-ImIm-(R)$^{(R)\text{-}CBI}$γ-PyPy-β-PyPy-β-Dp was recovered as a white powder upon lyophilization of the appropriate fraction (0.8 mg, 457 nmoles, 12.3 % recovery). UV (H_2O) λ_{max} (ε), 312nm, (73, 854); MALDI-TOF-MS (monoisotopic) [M+H] 1781.73 (calculated 1780.76 for $C_{83}H_{101}ClN_{28}O_{16}$).

ImImPyPy-β-PyPyPyPy-β-TA (3). ImImPyPy-β-PyPyPyPy-β-Pam resin was synthesized in a stepwise fashion by Boc-chemistry manual solid phase protocols. A sample of resin was treated with neat 3,3'-diamino-N-methyldipropylamine (triamine-TA) (2 ml), heated (55 °C, 24 hours) and purified by reversed phase HPLC. ImImPyPy-β-PyPyPyPy-β-TA was recovered as a white powder upon lyophilization of the appropriate fraction (14.2 mg, 11.4 μmoles, 14.8% recovery). UV (H_2O) λ_{max} (ε), 312 nm, (66, 600); MALDI-TOF-MS (monoisotopic) [M+H] 1251.68 (calculated 1250.60 for $C_{59}H_{74}N_{22}O_{10}$).

ImImPyPy-β-PyPyPyPy-β-TA-Glu-NHS (3-DSG). 3-DSG was prepared in an analogous fashion to **1-DSG** from z. ImImPyPy-β-PyPyPyPy-β-TA-Glu-NHS was recovered as a white powder upon lyophilization of the appropriate fraction. UV (H_2O) λ_{max} (ε), 312nm, (66, 600).

ImImPyPy-β-PyPyPyPy-β-TA-(S)-CBI (3S). 3S was prepared in an analogous fashion to **1R** from **3-DSG** and (S)-*seco*-CBI-β-ala. ImImPyPy-β-PyPyPyPy-β-TA-(S)-CBI was recovered as a white powder upon lyophilization of the appropriate fraction (0.74 mg, 446 nmoles, 9.4% recovery). UV (H_2O) λ_{max} (ε), 312nm, (73, 854); MALDI-TOF-MS (monoisotopic) [M+H] 1651.79 (calculated 1650.71 for $C_{80}H_{95}ClN_{24}O_{14}$).

DNA Reagents and Materials. Enzymes were purchased from Boehringer-Mannheim and used with their supplied buffers. Deoxyadenosine 5'-[γ-^{32}P] triphosphates were purchased from I.C.N. Sonicated, deproteinized calf thymus DNA was acquired from Pharmacia. RNase free water was obtained from USB and used for all reactions. All other reagents and materials were used as received. All DNA manipulations were performed according to standard protocols.

Construction of Plasmid DNA. The plasmids GFP1 and GFP2 were constructed by Dr. Matt Porteus. Concentration of the prepared plasmid was determined at 260 nm from the relationship of 1 OD unit=50 $\mu g\ mL^{-1}$ duplex DNA.

PCR Labeling to generate 5'-End-Labeled Restriction Fragments. Two 21 mer primers were synthesized for PCR amplification: primer A (top)5'-AAG-CAG-CAC-GAC-TTC-TTC-AAG-3' and primer B (bottom) 5'-CAG-GAT-GTT-GCC-GTC-CTC-CTT-3'. To label the top strand, primer A was treated with T4 polynucleotide kinase and deoxyadenosine 5'-[γ-^{32}P] triphosphate as previously described. PCR reactions containing 60 pmol each primer, 10 µl PCR buffer (Boehringer-Mannheim), 3.7 µl template (0.003 µg/mL), 2 µl dNTP mix (each at 10 mM), 1 µl 100X BSA (New England Biolabs) and 83 µl water were heated at 70 °C for 5 minutes. Four units of Taq Polymerase were added (Boehringer-Mannheim). Thirty amplification cycles were performed, each cycle consisting of the following segments: 94 °C for 1 minute, 54 °C for 1 minute, and 72 °C for 1.5 minutes. Following the last cycle, 10 minutes of extension at 72 °C completed the reaction. The PCR products were gel purified as previously reported for 3'-end labeling protocols. To label the bottom strand, primer B was labeled.

Cleavage Reactions. All reactions were carried out in a volume of 50 µl. A polyamide stock solution or water (for reference lane) was added to an assay buffer of TE (pH7.5) and 20 kcpm of 3'- or 5'-radiolabeled DNA. The solutions were allowed to equilibrate for 12 hours or the appropriate time (for time course reactions) at 37 °C. The reactions were stopped with 60 µl of a solution containing NaOAc (600 mM), EDTA pH 8.0 (12.5 mM), calf thymus DNA (150 µM base pair), glycogen (0.8 mg mL^{-1}), and NaCl (2 M). Ethanol was added to remove unbound polyamide and precipitate the products. The reactions were resuspended in 20 µl of TE (pH 7.5) and cleavage was initiated by heating at 95 °C for 30 minutes. The cleavage products were precipitated with 150 µl ethanol and then resuspended in 100 mM trisborate-EDTA/80% formamide loading buffer, and denatured and loaded onto polyacrylamide gels as previously reported. The gels were quantitated by the use of storage phosphor technology. Yield or efficiency of alkylation was determined as the ratio between the volume integration assigned to the products and the sum of the volumes of all the products in the lane.

Acknowledgements

We are grateful to the National Institutes of Health for research support and the Ralph M. Parsons foundations for a predoctoral fellowship to A.Y.C. We thank Dr. Matt Porteus of the Baltimore group at Caltech and Mr. Nicholas Wurtz of the Dervan group for their collaborative efforts on this project. We thank G.M. Hathaway and the Caltech Protein/Peptide Microanalytical Laboratory for MALDI-TOF mass spectrometry.

References

1. Dickinson, L. A. et al. Inhibition of RNA polymerase II transcription in human cells by synthetic DNA-binding ligands. *Proc. Natl. Acad. Sci. USA* **95**, 12890-12895 (1998).

2. Mapp, A. K., Ansari, A. Z., Ptashne, M. & Dervan, P. B. Activation of gene expression by small molecule transcription factors. *Proc. Natl. Acad. Sci. USA* **97**, 3930-3935 (2000).

3. Chang, A. Y. & Dervan, P. B. Strand selective cleavage of DNA by diastereomers of hairpin polyamide-seco-CBI conjugates. *J. Am. Chem. Soc.* **122**, 4856-4864 (2000).

4. Jasin, M. Genetic manipulation of genomes with rare-cutting endonucleases. *Trends Genet.* **12**, 224-228 (1996).

5. Giovannangeli, C. & Helene, C. Triplex technology takes off. *Nat. Biotechnol.* **18**, 1245-1246 (2000).

6. Vasquez, K. M., Narayanan, L. & Glazer, P. M. Specific mutations induced by triplex-forming oligonucleotides in mice. *Science* **290**, 530-533 (2000).

7. White, S., Szewczyk, J. W., Turner, J. M., Baird, E. E. & Dervan, P. B. Recognition of the four Watson-Crick base pairs in the DNA minor groove by synthetic ligands. *Nature* **391**, 468-471 (1998).

8. Gottesfeld, J. M., Neely, L., Trauger, J. W., Baird, E. E. & Dervan, P. B. Regulation of gene expression by small molecules. *Nature* **387**, 202-205 (1997).

9. Baird, E. E. & Dervan, P. B. Solid phase synthesis of polyamides containing imidazole and pyrrole amino acids. *J. Am. Chem. Soc.* **118**, 6141-6146 (1996).

10. *Swalley*, S. E., Baird, E. E. & Dervan, P. B. A pyrrole-imidazole polyamide motif for recognition of eleven base pair sequences in the minor groove of DNA. *Chem.-Eur. J.* **3**, 1600-1607 (1997).

11. Herman, D. M., Baird, E. E. & Dervan, P. B. Stereochemical control of the DNA binding affinity, sequence specificity, and orientation preference of chiral hairpin polyamides in the minor groove. *J. Am. Chem. Soc.* **120**, 1382-1391 (1998).

12. Swalley, S. E., Baird, E. E. & Dervan, P. B. Recognition of a 5'-(A,T)GGG(A,T)(2)-3' sequence in the minor groove of DNA by an eight-ring hairpin polyamide. *J. Am. Chem. Soc.* **118**, 8198-8206 (1996).

13. Wurtz, N. R. & Dervan, P. B. Sequence Specific Alkylation of DNA by Pyrrole-Imidazole Hairpin Polyamide Conjugates. *Chem. Biol.* (1999).

APPENDIX ONE

Polyamide-*seco*-CBI Conjugates to Study COX-2 and HIV-1

Abstract

Described are the synthesis and characterization for *seco*-CBI polyamide conjugates to study inhibition of transcription elongation of COX-2 and to probe the cellular activity of transcription inhibition in HIV-1. Four compounds were studied for COX-2: ImPyPy-(R)[ImPyIm-(R)^{H_2N}γ–PyPyPyδ]^{HN}γPyPyPy-β-Dp (**1**), Im-β-ImPyPyPy-γ^{H_2N}-ImPyPyPy-β-Py-β-Dp (**2**), ImPy-β-ImIm-γ^{H_2N}-PyPy-β-ImPy-β-Dp (**3**), ImPy-β-ImPy-γ^{H_2N}-PyPy-β-ImPy-β-Dp (**4**). Two were studied for HIV-1: ImPy-β-ImPy-γ$^{(S)\text{-}CBI}$-ImPy-β-ImPy-β-Dp (**5**), and ImIm-β-ImIm-γ$^{(S)\text{-}CBI}$-PyPy-β-PyPy-β-Dp (**6**). Examination of the alkylation patterns of the polyamides described offers insights into the design of future generations of this class of molecules. Further testing of these compounds in cellular assays will be performed shortly. We anticipate that those results will help us address new design considerations that may be necessary for future compounds in this series.

Polyamides for Inhibition of COX-2 Transcription Elongation

Background

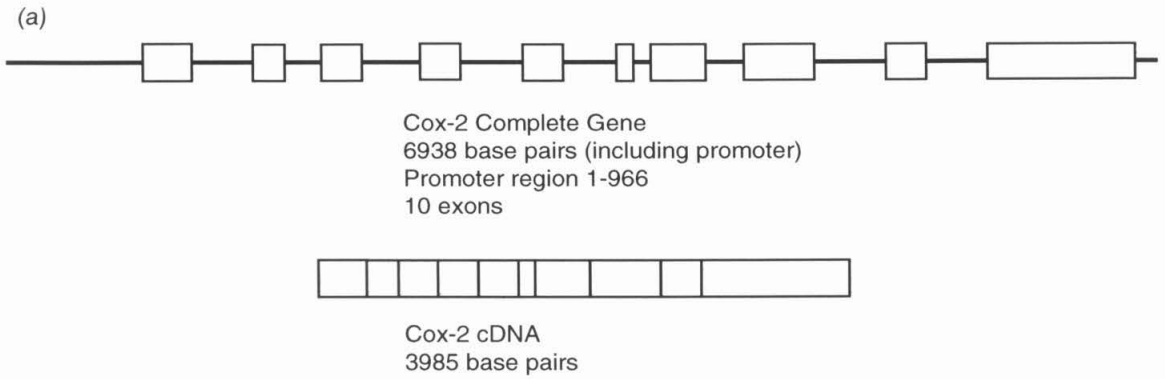
In an inflammatory response, cyclooxygenase (COX) catalyzes the first committed step in arachidonic acid metabolism, eventually resulting in the formation of prostaglandins. There are two distinct COX enzymes, COX-1 and COX-2. COX-1 is expressed in most tissues and is thought to release prostaglandins for the purposes of cellular housekeeping for maintaining organ and tissue homeostasis. COX-2 is expressed only in certain cells, such as fibroblasts and macrophages. It can be rapidly induced by cytokines or other mitogens. Inhibition of COX-2 relieves symptoms of inflammation, pain, and fever, while inhibition of COX-1 leads to side effects such as nausea and stomach pains.

NSAIDs (nonsteroidal anti-inflammatory drugs) are inhibitors of COX, and include aspirin, ibuprofen, and naproxen. One problem with NSAIDs is a lack of specificity between the two COX enzymes. Often they inhibit both enzymes, resulting in both the alleviation of some symptoms and the onset of others. The crystal structures of COX-1 and COX-2 show that the two enzymes are extrememly similar in overall structure, with an r.m.s deviation of 0.9 Å for Cα atoms.[1] The most significant difference in the structures is within the binding pocket. An isoleucine in COX-1 is replaced by a valine residue in COX-2. Many of the newer NSAIDs try to take advantage of this larger binding pocket in COX-2 by making bulkier substituents which are sterically unable to fit in the COX-1 binding site.[1,2]

Experimental Design

Prof. Harvey Herschman of UCLA was one of the early discoverers of COX-2,[3] and a collaboration with the Dervan group was started to see if Py-Im polyamides would be able to inhibit the induction of COX-2 by interfering with required transcription factors. We were also interested in looking at Py-Im polyamide conjugates with alkylating agents to see if transcription elongation of COX-2 could be inhibited by forming a covalent adduct with the DNA. Mr. Nicholas Wurtz from the Dervan group had already established a system of polyamide-nitrogen mustard conjugates that were promising as sequence specific DNA alkylating agents.[4] Together, we designed a series of compounds to incorporate both the *seco*-CBI conjugates and the nitrogen mustard conjugates.

The COX-2 gene is 6938 base pairs long, including the promoter (966 base pairs). It contains ten exons, which comprise the 3985 base pairs of the COX-2 cDNA. From the complete gene sequence, we found four target sequences which occurred anywhere from twice to thirteen times within the coding region of COX-2, and also had an adenine proximal to the binding site (Figure 1). The Herschman group supplied us with a plasmid of the COX-2 cDNA, and we used fragments from the cDNA to determine the binding affinities of the unmodified parent polyamides. To minimize the number of fragments we had to label, we selected two fragments to examine, 198-496 (numbering from cDNA sequence) and 500-772. Between the two fragments, they contain the binding sites of all four parent polyamides (Figure 2).



(b)

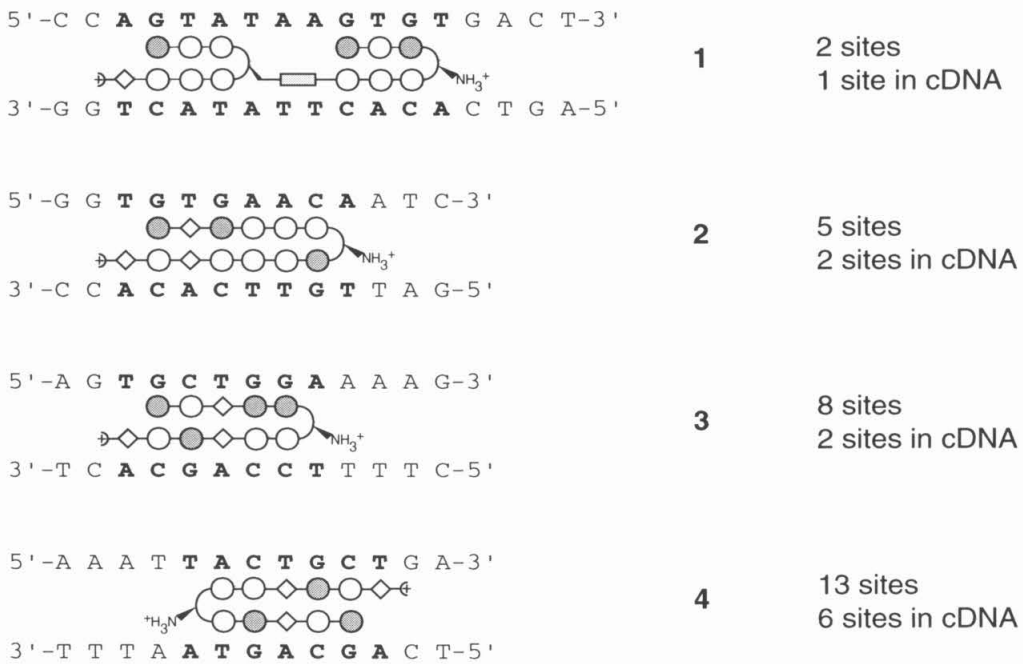


Figure 1. (a) Pictorial view of COX-2 gene. (b) Parent polyamides selected for this study, with number of sites in the COX-2 gene indicated.

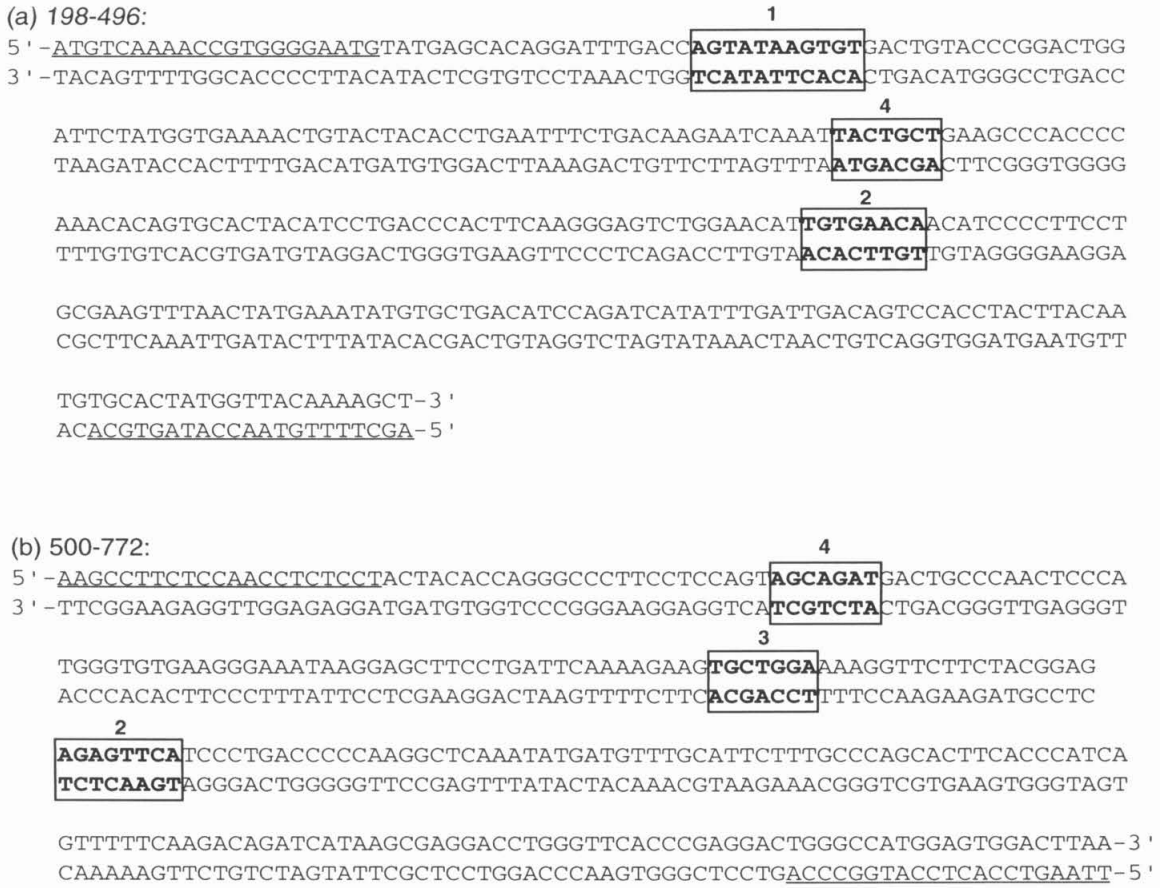


Figure 2. Fragments of COX-2 cDNA used for this study. (a) 198-496, (b) 500-772. Boxes indicate the binding sites of the corresponding polyamide. Underlined sequence corresponds to the primers used for labeling studies.

Results and Discussion

Synthesis and Footprinting of Parent Polyamides

Polyamides **1**-**4** were synthesized according to established solid phase synthesis protocols (Figure 3).[5] Nicholas Wurtz performed the footprinting on compounds **1** and **3**, and I performed footprinting on **2** and **4**.

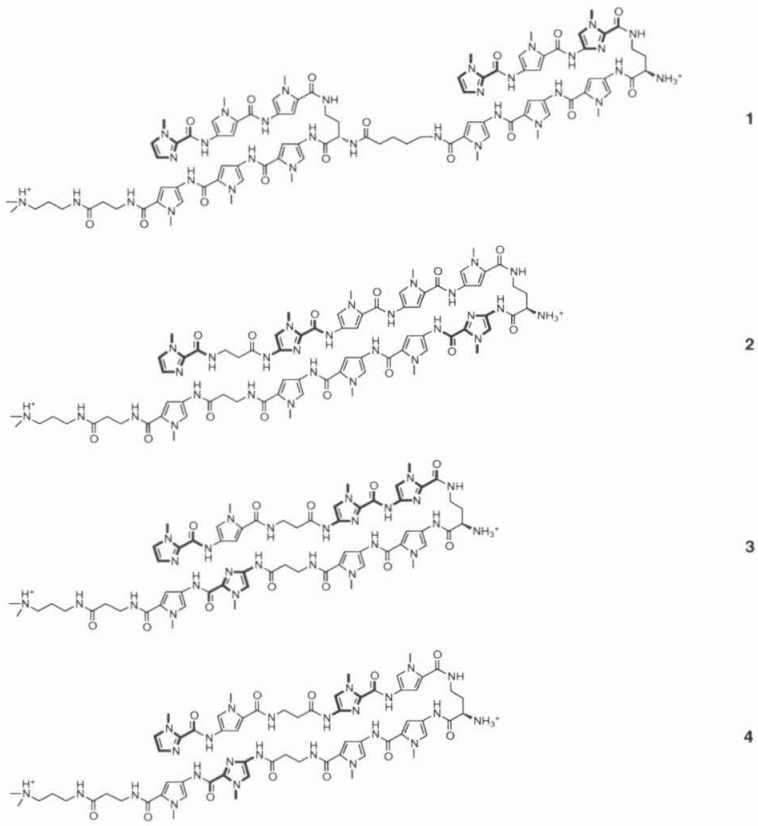


Figure 3. Structures of the parent polyamides used in the COX-2 study.

Tandem polyamide **1** ImPyPy-(R)[ImPyIm-(R)^{H_2N}γ–PyPyPyδ]^{HN}γPyPyPy-β-Dp, when measured on fragment 198-496 had an affinity of 4.0×10^9 at the site 5'-AGTATAAGTGT-3.' We had anticipated that this compound would have a higher binding affinity, but were unsure of the generality of the tandem motif. When compared to the tandem ImPyPy-(R)[ImPyPy-(R)^{H_2N}γ–PyPyPyδ]^{HN}γPyPyPy-β-Dp which bound the 11 base pair sequence 5'-TGTTATTGTTA-3' with a K_a of $\geq 1 \times 10^{12}\,M^{-1}$, the affinity of polyamide **1** is lower than expected.[6] However, the binding affinity should be adequate for our purposes.

Polyamide **2** Im-β-ImPyPyPy-γ^{H_2N}-ImPyPyPy-β-Py-β-Dp when measured on fragment 198-496 had an affinity of $2.3 \times 10^{10}\,M^{-1}$ at the sequence 5'-TGTGAACA-3.

This compares to the K_a of $\geq 2.4 \times 10^{10}\,M^{-1}$, that was reported for the compound Im-β-ImPyPyPy-γ-ImPyPyPy-β-Py-β-Dp at the same sequence.[7] Polyamides **3** and **4** had binding affinities of $1.8 \times 10^{10}\,M^{-1}$ and $1.1 \times 10^{10}\,M^{-1}$, respectively, at the targeted match sites of 5'-TGCTGGA-3' for polyamide **3** and 5'-TACTGCT-3' for polyamide **4**.

Synthesis and Thermal Cleavage Reactions of Polyamide Conjugates

Because the footprinting experiments revealed that the polyamides bound to the designed target sequences, the polyamide *seco*-CBI conjugates **1S** and **3S** were synthesized as previously reported (Figure 4).[8]

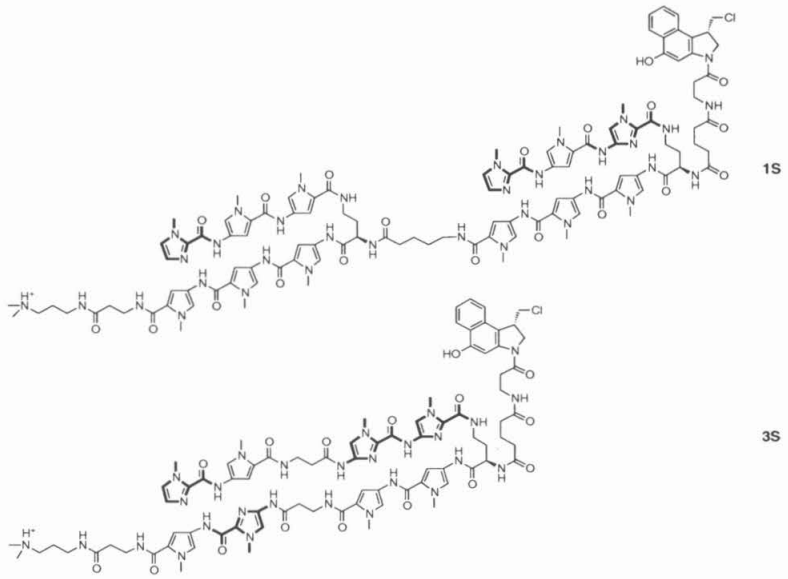


Figure 4. Chemical structures of the polyamide conjugates used for the COX-2 study.

Thermal cleavage reactions were performed for compound **1S** on fragment 198-496 (Figure 5). Surprisingly, the alkylation gels showed that instead of alkylating proximal to the match site 5'-AGTATAAGTGT-3', a strong alkylation site is observed proximal to the mismatch site 5'-TTCTATGGTGA-3' (mismatch underlined). The sequence requirements of the valeric acid linker have yet to be determined, so the G in the valeric acid linker region may or may not be a mismatch. At a minimum, this site is a

double base pair mismatch, and possibly a triple base pair mismatch. Alkylation at this site has a yield of 79% at 10 nM concentrations. No alkylation is observed at the adenine proximal to the match site. In addition to alkylating at a mismatch site, alkylation occurs three base pairs removed from the binding site of the polyamide, not the anticipated two base pairs. The reasons for this anomalous alkylation pattern shown for **1S** are unclear. One possibility is that the adenine proximal to the intended match site of **1S** is flanked by two G-C base pairs. The mixed sequence is not optimal for alkylation by CBI. However, the site that is alkylated by **1S** is flanked by an A-T tract, which is known to be a consensus sequence for the duocarmycins. This may be an instance where k_{off} is slow at the match site, but k_{alk} is slower, resulting in no alkylation. Meanwhile, k_{off} may be fast at the mismatch site, but k_{alk} is faster, yielding a strong cleavage site.

Polyamide **3S** was analyzed on fragment 500-772, and shows alkylation proximal to the expected match site 5'-TGCTGGA-3' (Figure 6). At 300 pM concentrations, the cleavage yield at this site is 83%. Again, an interesting observation is that the cleavage occurs at the adenine three base pairs removed from the binding site instead of two. The alkylated adenine is at the 3' side of a short A-T tract. It is known that A-T tracts bend toward the minor groove.[9] Perhaps the bend in the DNA causes the unexpected 3' shift in alkylation.

Future Work

Due to problems generating reproducible and reliable results of inhibition of transcription initiation, work on this project has slowed. Once assays are available that can detect changes in protein or mRNA levels reproducibly, it will be interesting to see what polyamide-alkylator conjugates do when targeted to the coding region. Even

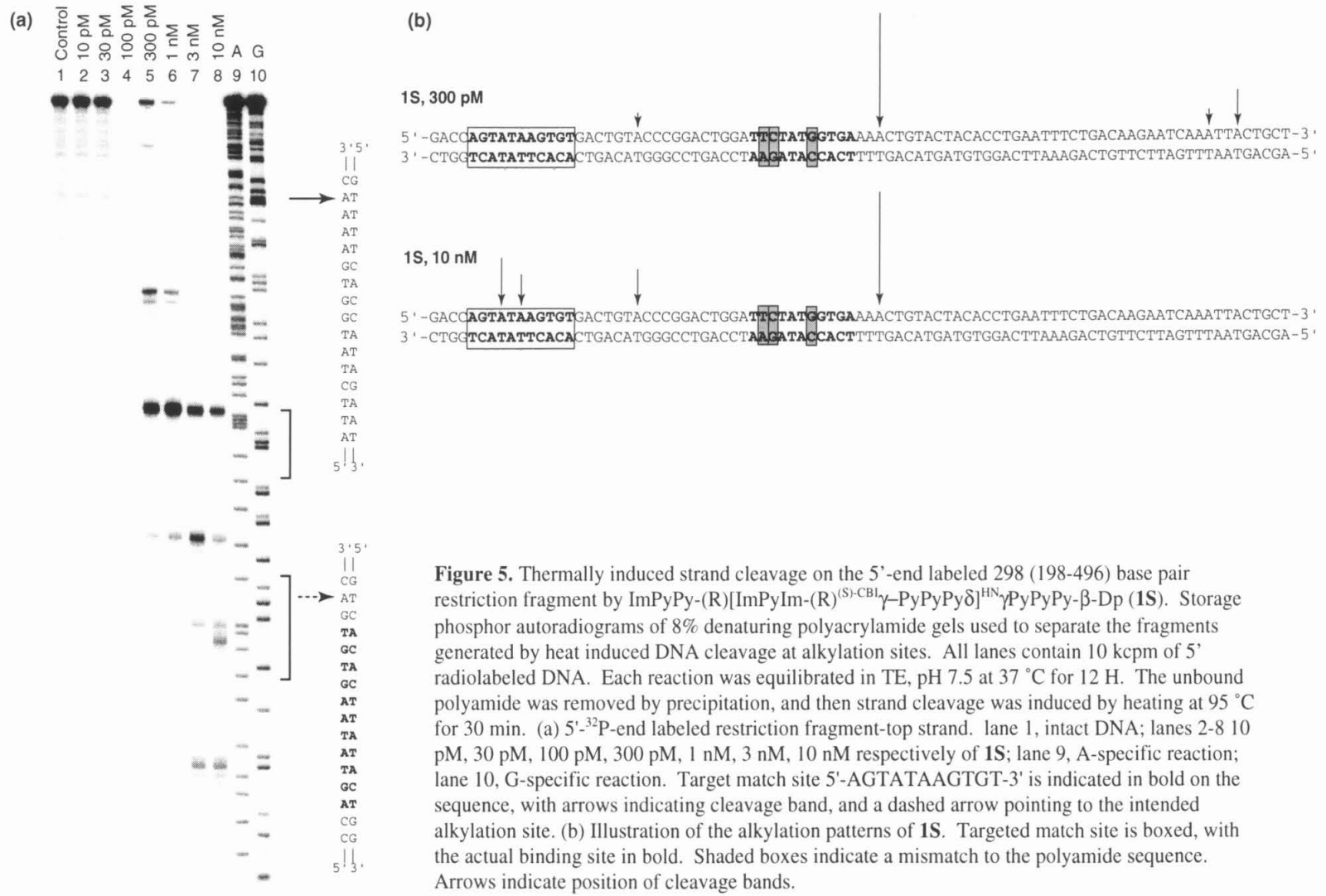


Figure 5. Thermally induced strand cleavage on the 5'-end labeled 298 (198-496) base pair restriction fragment by ImPyPy-(R)[ImPyIm-(R)$^{(S)\text{-}CBI}$γ–PyPyPyδ]^{HN}γPyPyPy-β-Dp (**1S**). Storage phosphor autoradiograms of 8% denaturing polyacrylamide gels used to separate the fragments generated by heat induced DNA cleavage at alkylation sites. All lanes contain 10 kcpm of 5' radiolabeled DNA. Each reaction was equilibrated in TE, pH 7.5 at 37 °C for 12 H. The unbound polyamide was removed by precipitation, and then strand cleavage was induced by heating at 95 °C for 30 min. (a) 5'-^{32}P-end labeled restriction fragment-top strand. lane 1, intact DNA; lanes 2-8 10 pM, 30 pM, 100 pM, 300 pM, 1 nM, 3 nM, 10 nM respectively of **1S**; lane 9, A-specific reaction; lane 10, G-specific reaction. Target match site 5'-AGTATAAGTGT-3' is indicated in bold on the sequence, with arrows indicating cleavage band, and a dashed arrow pointing to the intended alkylation site. (b) Illustration of the alkylation patterns of **1S**. Targeted match site is boxed, with the actual binding site in bold. Shaded boxes indicate a mismatch to the polyamide sequence. Arrows indicate position of cleavage bands.

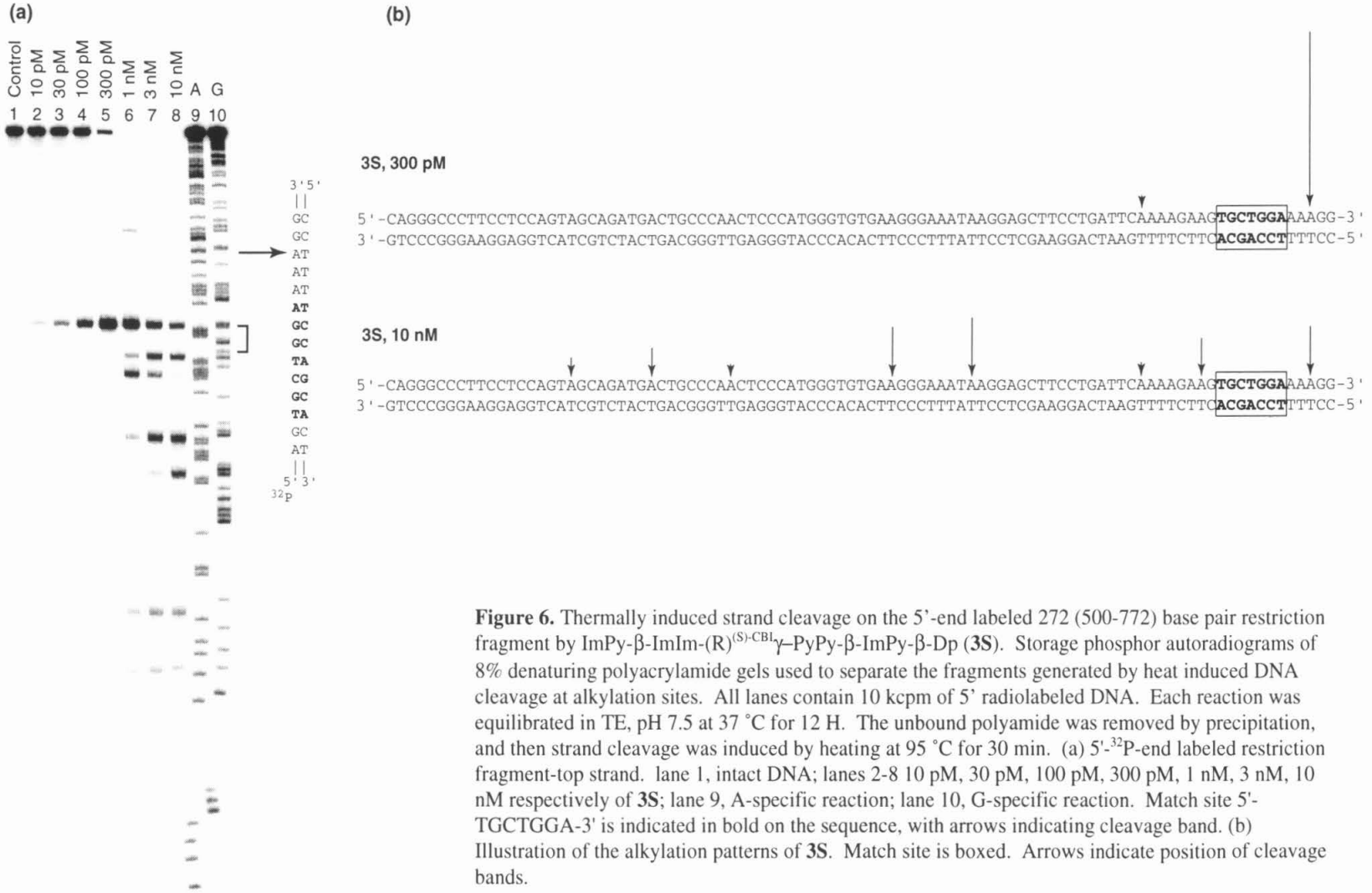


Figure 6. Thermally induced strand cleavage on the 5'-end labeled 272 (500-772) base pair restriction fragment by ImPy-β-ImIm-(R)$^{(S)\text{-}CBI}$γ–PyPy-β-ImPy-β-Dp (**3S**). Storage phosphor autoradiograms of 8% denaturing polyacrylamide gels used to separate the fragments generated by heat induced DNA cleavage at alkylation sites. All lanes contain 10 kcpm of 5' radiolabeled DNA. Each reaction was equilibrated in TE, pH 7.5 at 37 ˚C for 12 H. The unbound polyamide was removed by precipitation, and then strand cleavage was induced by heating at 95 ˚C for 30 min. (a) 5'-^{32}P-end labeled restriction fragment-top strand. lane 1, intact DNA; lanes 2-8 10 pM, 30 pM, 100 pM, 300 pM, 1 nM, 3 nM, 10 nM respectively of **3S**; lane 9, A-specific reaction; lane 10, G-specific reaction. Match site 5'-TGCTGGA-3' is indicated in bold on the sequence, with arrows indicating cleavage band. (b) Illustration of the alkylation patterns of **3S**. Match site is boxed. Arrows indicate position of cleavage bands.

without the biological assay results, it is worthy of note that for the two polyamide motifs in this study, both generally followed previous patterns of *seco*-CBI conjugates, and had high cleavage yields.

Polyamide-Alkylator Conjugates that Target the Promoter Region of HIV-1

Background

Described previously were successful efforts to use Py-Im polyamides to inhibit binding of essential transcription factors to inhibit transcription of HIV-1. Two of the compounds used in that study were **5** and **6** (Figure 7a).[10] Polyamide **5**, ImPy-β-ImPy-γ-ImPy-β-ImPy-β-Dp, binds to the sequence 5'-WGCWGCW-3' with a K_d of 0.05 nM. There are two match sites for polyamide **5** flanking the binding site of TATA Binding Protein (TBP) in the HIV-1 promoter. Polyamide **6**, ImIm-β-ImIm-γ-PyPy-β-PyPy-β-Dp serves as a control compound and is a double base pair mismatch for the sequence 5'-WGCWGCW-3', and binds that site with a K_d of 5.0 nM. This study reported that polyamide **5** could inhibit TBP binding *in vitro*, while polyamide **6** could not. Inhibition of *in vitro* transcription by 50% was observed at 60 nM concentrations of polyamide **5**, while no inhibition was observed for polyamide **6**. Cell culture experiments indicated that polyamide **5** inhibited HIV-1 viral replication by 80% at 1 μM concentrations, and polyamide **6** had no effect on viral replication.

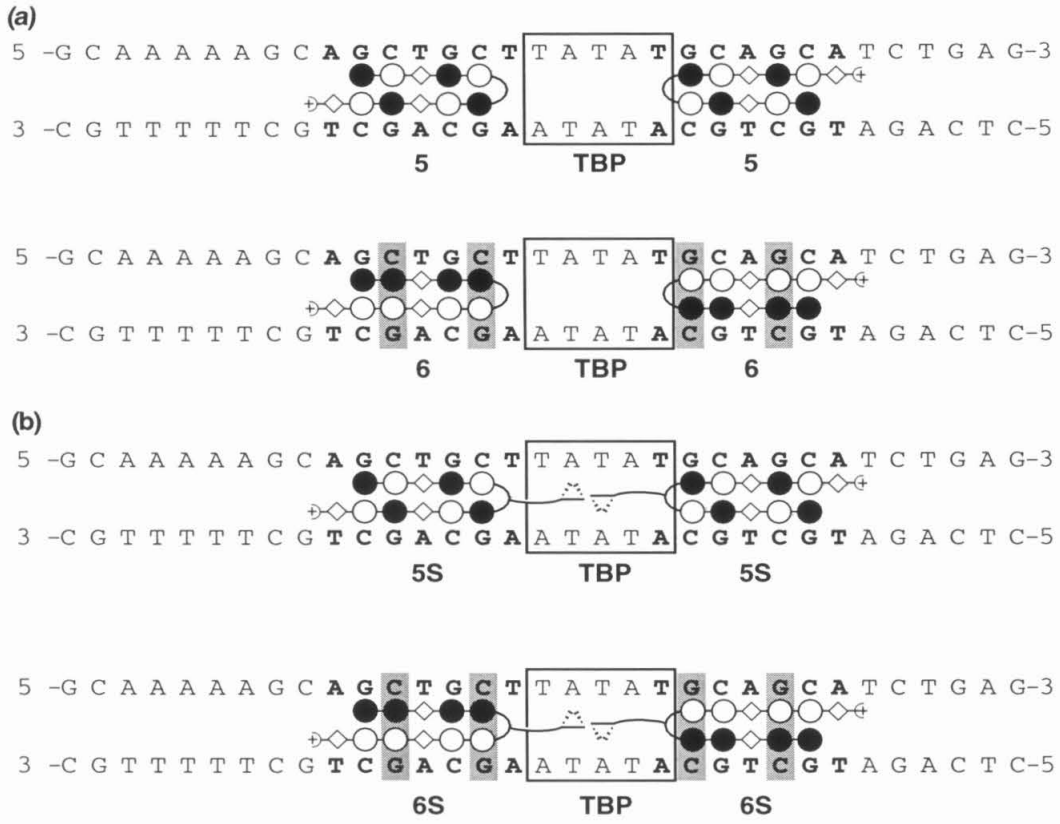


Figure 7. (a) Ball and stick model of the polyamides used for inhibition of HIV-1 transcription initiation which inhibit binding of TBP. (b) Proposed *seco*-CBI conjugates to target to HIV-1 promoter.

A derivative of polyamide **5** modified with a nitrogen mustard has already been extensively characterized in the Dervan group by Nicholas Wurtz.[4] It was found that the nitrogen mustard derivatized polyamide was able to alkylate the HIV-LTR promoter with good specificity and efficiency. Joel Gottesfeld of The Scripps Research Institute was a collaborator on the transcription inhibition studies of polyamides **5** and **6**. He was interested in studying the nitrogen mustard analogues to confirm that polyamides are able to get to the nucleus and target their match sequence. By analyzing the cellular genomic DNA, he wanted to analyze the cleaved DNA resulting from the alkylation by the nitrogen mustard. To that end, we also designed *seco*-CBI polyamide conjugates of

polyamides ***5*** and ***6*** to compare to the results of the nitrogen mustard analogues (Figure 7b).

Synthesis and Thermal Cleavage Assays

Polyamides **5S** and **6S** were synthesized as previously described (Figure 8).[5,8] They were analyzed on a 241 base pair fragment containing the HIV-LTR promoter (Figure 9).

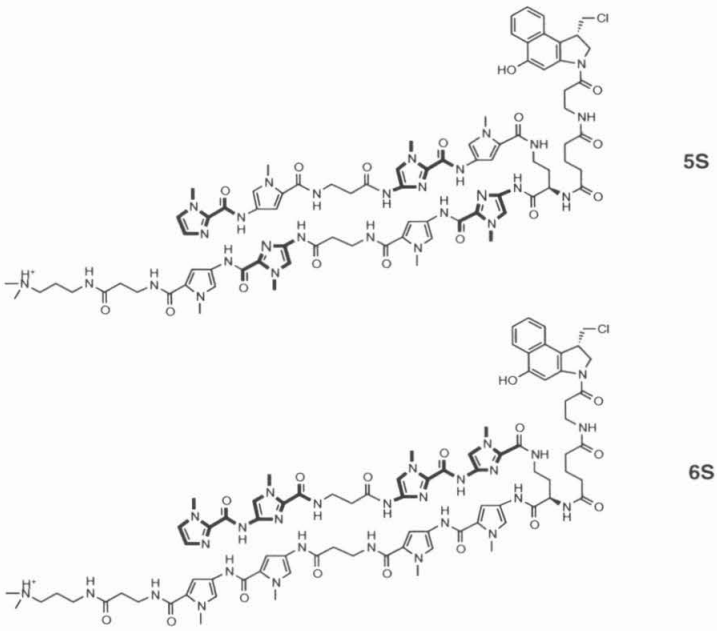


Figure 8. Chemical structures of the *seco*-CBI polyamide conjugates used for the HIV-1 study.

pHIV-LTR restriction fragment (241 bp)

```
5'-AATTCGAGCTCGGTACCCGGTAACCAGAGAGACCCAGTACAGGCAAAAAGCAGCTGCTTATATGCAGCATCTGAGGGACG
3'-TTAAGCTCGAGCCATGGGCCATTGGTCTCTCTGGGTCATGTCCGTTTTTCGTCGACGAATATACGTCGTAGACACCCTGC

CCACTCCCCAGTCCCGCCCCAGGCCACGCCTCCCTGGAAAGTCCCCAGCGGAAAGTCCCTTGTAGAAAGCTCGATGTCAGCAG
GGTGAGGGGTCAGGGCGGGGTCCGGTGCGGAGGGACCTTTCAGGGGTCGCCTTTCTGGGAACATCTTTCGAGATACAGTCGTC

TCTTTGTAGTACTCCGGATGCAGCTCTCGGGCCACGTGCTGAAATGCTAGGCGGCTGTCAATCGACCTGCAGGCATGCA-3'
AGAAACATCATGAGGCCTACGACGAGAGCCCGGTGCACGACTTTACGATCCGCCGACAGTTAGCTGGACGTCCGTACGT-5'
```

Figure 9. HIV-LTR restriction fragment used in this study. Polyamide binding sites are underlined and TBP binding site is indicated in bold.

Polyamide **5S** shows alkylation proximal to the binding sites flanking the TATA box (Figure 10). At the match site to the right of the TATA box, 5'-TGCAGCA-3', polyamide **5S** shows alkylation at two adenines. It is a bit unusual that alkylation is seen at adenines so far from the polyamide binding site, but may be explained because the TATA box is known to be bent toward the minor groove.[9] The cleavage yield at this site is 45% at 1 nM concentrations of **5S**. The binding site to the left of the TATA box is actually two overlapping match sites, 5'-AGCAGCTGCT-3'. Here a high yielding cleavage site (62% at 10 nM) is observed along an A-T tract. Another alkylation site at the TATA box is observed for another binding mode of the polyamide. This site has a lower cleavage yield of 12% at 3 nM concentrations.

Despite being designed as a mismatch, polyamide **6S** cleaves the same fragment at almost identical adenines with comparable or higher cleavage yields (Figure 11). Again, this may be a situation where the equilibration of the polyamide off of a mismatch site (k_{off}) is slower than the alkylation reaction of *seco*-CBI at N3 of adenine (k_{alk}). The A-T tracts that flank the polyamide binding sites are known to be preferred alkylation sites for the duocarmycins, so this is a reasonable possibility.

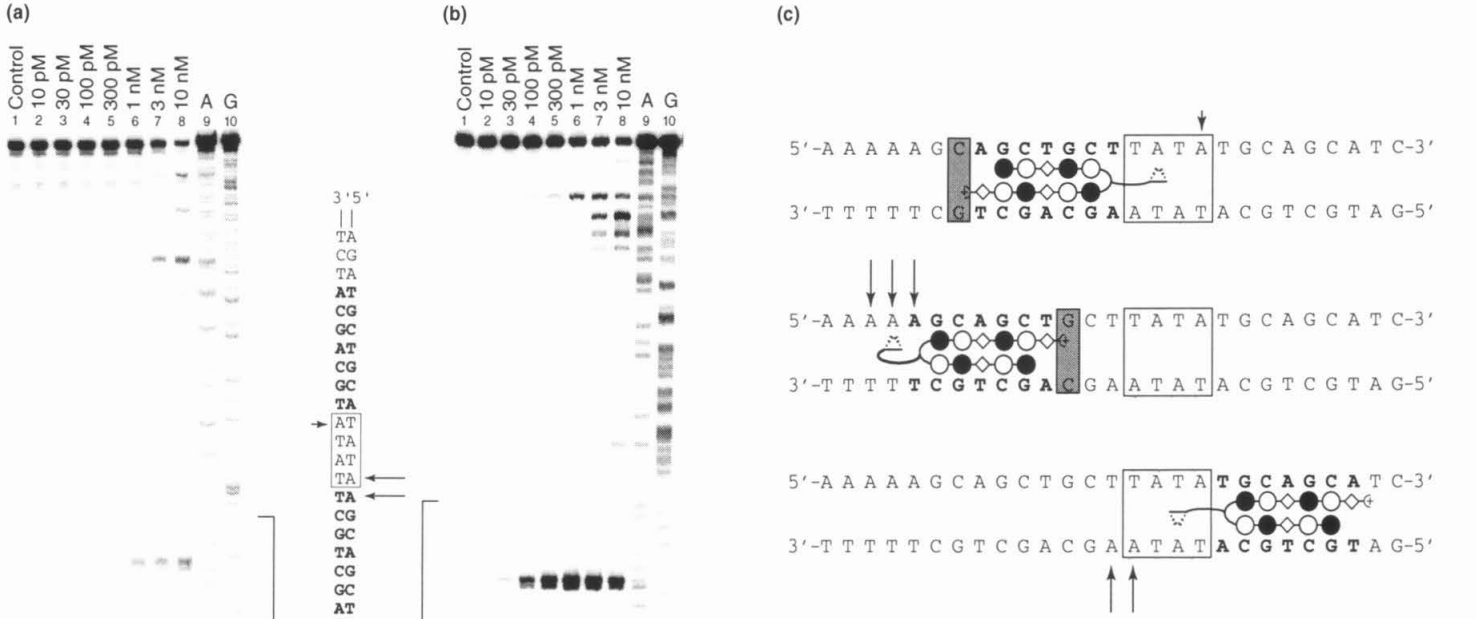


Figure 10. Thermally induced strand cleavage on the 5'-end labeled and 3'-end labeled 241 base pair restriction fragment by ImPy-β-ImPy-(R)$^{(S)\text{-}CBI}$γ-ImPy-β-ImPy-β-Dp (**5S**). Storage phosphor autoradiograms of 8% denaturing polyacrylamide gels used to separate the fragments generated by heat induced DNA cleavage at alkylation sites. All lanes contain 10 kcpm of either 5' or 3' radiolabeled DNA. Each reaction was equilibrated in TE, pH 7.5 at 37 °C for 12 H. The unbound polyamide was removed by precipitation, and then strand cleavage was induced by heating at 95 °C for 30 min. (a) 5'-^{32}P-end labeled restriction fragment-top strand. (b) 5'-^{32}P-end labeled restriction fragment-bottom strand. (a-b) lane 1, intact DNA; lanes 2-8 10 pM, 30 pM, 100 pM, 300 pM, 1 nM, 3 nM, 10 nM respectively of **5S**; lane 9, A-specific reaction; lane 10, G-specific reaction. Match site 5'-TGCTGGA-3' is indicated in bold on the sequence, with arrows indicating cleavage band. (b) Illustration of the alkylation patterns of **5S**. Match site is in bold, and the TBP site is boxed. Shaded boxes indicate a mismatch to the polyamide sequence. Arrows indicate position of cleavage bands.

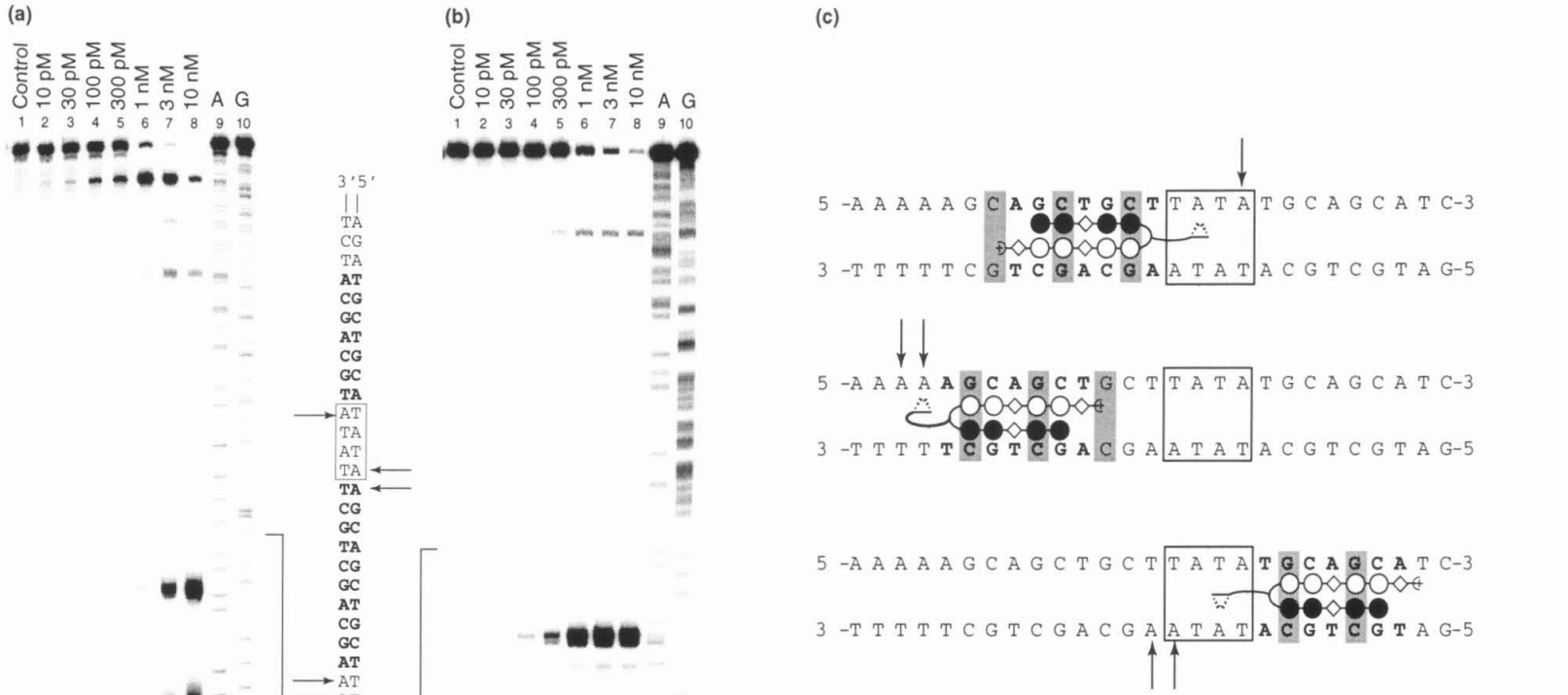


Figure 11. Thermally induced strand cleavage on the 5'-end labeled and 3'-end labeled 241 base pair restriction fragment by ImIm-β-ImIm-(R)$^{(S)\text{-}CBI}$γ-PyPy-β-PyPy-β-Dp (**6S**). Storage phosphor autoradiograms of 8% denaturing polyacrylamide gels used to separate the fragments generated by heat induced DNA cleavage at alkylation sites. All lanes contain 10 kcpm of either 5' or 3' radiolabeled DNA. Each reaction was equilibrated in TE, pH 7.5 at 37 °C for 12 H. The unbound polyamide was removed by precipitation, and then strand cleavage was induced by heating at 95 °C for 30 min. (a) 5'-^{32}P-end labeled restriction fragment-top strand. (b) 5'-^{32}P-end labeled restriction fragment-bottom strand. (a-b) lane 1, intact DNA; lanes 2-8 10 pM, 30 pM, 100 pM, 300 pM, 1 nM, 3 nM, 10 nM respectively of **6S**; lane 9, A-specific reaction; lane 10, G-specific reaction. Match site 5'-TGCTGGA-3' is indicated in bold on the sequence, with arrows indicating cleavage band. (b) Illustration of the alkylation patterns of **6S**. Match site is in bold, and the TBP site is boxed. Shaded boxes indicate a mismatch to the polyamide sequence. Arrows indicate position of cleavage bands.

Comparison with Nitrogen Mustard Polyamide Conjugates

Figure 12 compares the alkylation sites seen with the polyamide *seco*-CBI conjugates and the nitrogen mustard conjugates. The nitrogen mustard analogue of polyamide **6** showed no alkylation at all on this fragment. Comparing **5S** and its nitrogen mustard counterpart, at the binding site on the left side of the TATA box (Figure 12a), the nitrogen mustard analogue alkylates preferentially at the TATA box, and shows no alkylation at the A-T tract. Meanwhile, **5S** shows the majority of cleavage at the A-T tract. At the binding site to the right side of the TATA box (Figure 12b), polyamide **5S** displays DNA cleavage at the TATA box, as does the nitrogen mustard analogue. The nitrogen mustard analogue also shows alkylation at an adenine on the opposite side of the TATA box for the other polyamide binding mode.

The differences between the nitrogen mustard polyamide conjugates and the *seco*-CBI conjugates are noteworthy. As DNA alkylating agents, nitrogen mustards prefer to alkylate in the major groove at N7 of guanine, but will react with N3 of adenine in the minor groove when directed by a minor groove binding ligand.[11,12] Conversely, duocarmycins target the N3 of adenine, and react preferentially there. The differences observed in the alkylation patterns most likely result from the different reactivity of the alkylating agents. The results indicate that at particularly reactive sites for *seco*-CBI, alkylation occurs more quickly than equilibration from a mismatch site. Meanwhile, for the nitrogen mustard conjugates, because the reactivity is slower at N3 of adenine, it waits for the polyamide to equilibrate to a match site and then alkylates DNA if an adenine is available. So, decreased rate of alkylation actually allows for increased specificity of alkylation.

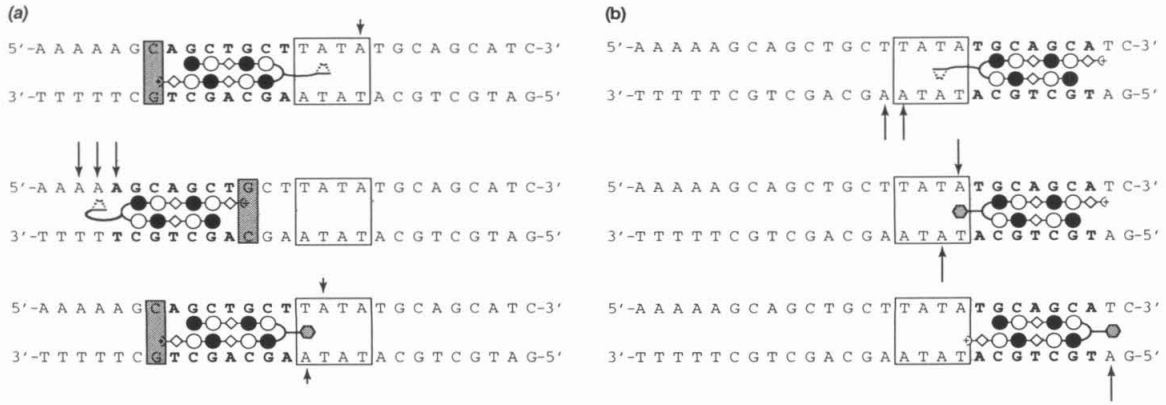


Figure 12. Ball and stick models that compare the alkylation of *seco*-CBI polyamide conjugates with nitrogen mustard conjugates. Symbols are as described before, with the shaded hexagon representing the nitrogen mustard.

Conclusions

Currently, these *seco*-CBI conjugates have been sent to the Gottesfeld group and we look forward to seeing how these molecules react in cells.

The investigation of *seco*-CBI conjugated to polyamides of different sequence and motif suggests additional factors that influence reactivity of such compounds. When designing such compounds to address a particular application, the sequence specificity and affinity of the Py-Im polyamide will always be of primary importance. However, the kinetics of polyamide binding also affect the specificity of such conjugates, because of their fast on rates to DNA, and slow off rates.[13] The sequence context of the binding site in question should not be overlooked. In the case of *seco*-CBI conjugates, it appears that flanking sequences and the reactivity of target alkylation sites greatly affect the efficiency and predictable nature of DNA alkylation. Those things can be taken into consideration for future generations of these compounds.

Experimental

Materials

^{1}H NMR spectra were recorded on a General Electric-QE NMR spectrometer at 300 MHz with chemical shifts reported in parts per million relative to residual solvent. UV spectra were measured in water on a Hewlett-Packard Model 8452A diode array spectrophotometer. Matrix-assisted, laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF) was performed at the Protein and Peptide Microanalytical Facility at the California Institute of Technology. Preparatory reversed phase HPLC was performed on a Beckman HPLC with a Waters DeltaPak 25 × 100 mm, 300 Å C18 column equipped with a guard, 0.1% (wt/v) TFA, 0.25% acetonitrile/min.

Synthesis of CBI-polyamide conjugates

ImPyPy-(R)[ImPyIm-(R)^{H_2N}γ–PyPyPyδ]^{HN}γPyPyPy-β-Dp (1). ImPyPy-(R)[ImPyIm-(R)^{H_2N}γ–PyPyPyδ]^{HN}γPyPyPy-β-Pam resin was synthesized in a stepwise fashion by Boc-chemistry manual solid phase protocols by Nicholas Wurtz. A sample of resin was treated with neat (dimethylamino)-propylamine (2 ml), heated (55 °C, 24 hours) and purified by reversed phase HPLC. ImPyPy-(R)[ImPyIm-(R)^{H_2N}γ–PyPyPyδ]^{HN}γPyPyPy-β-Dp was recovered as a white powder upon lyophilization of the appropriate fraction. UV (H_2O) λ_{max} (ε), 312 nm, (99, 900); MALDI-TOF-MS (monoisotopic) [M+H] 1910.97 (calculated 1909.97 for $C_{91}H_{111}N_{33}O_{16}$).

ImPyPy-(R)[ImPyIm-(R)$^{Glu\text{-}NHS}$γ–PyPyPyδ]^{HN}γPyPyPy-β-Dp (1-DSG). 1-DSG was prepared by already published methods.[8] UV (H_2O) λ_{max} (ε), 312 nm, (99, 900); MALDI-TOF-MS (monoisotopic) [M+H] 2122.25 (calculated 2120.94 for $C_{99}H_{120}N_{34}O_{21}$).

***ImPyPy-(R)[ImPyIm-(R)$^{(S)\text{-}CBI}$γ–PyPyPyδ]^{HN}γPyPyPy-β-Dp* (1S).** **1S** was prepared by already published methods.[8] UV (H_2O) λ_{max} (ε), 312 nm, (107, 154); MALDI-TOF-MS (monoisotopic) [M+H] 2311.12 (calculated 2310.01 for $C_{111}H_{132}ClN_{35}O_{20}$).

Im-β-ImPyPyPy-(R)^{H_2N}γ-ImPyPyPy-β-Py-β-Dp (2). **2** was prepared in an analogous fashion to **1**. UV (H_2O) λ_{max} (ε), 312 nm, (83, 250); MALDI-TOF-MS (monoisotopic) [M+H] 1624.88 (calculated 1623.75 for $C_{75}H_{93}N_{29}O_{14}$).

ImPy-β-ImIm-(R)^{H_2N}γ-PyPy-β-ImPy-β-Dp (3). **3** was prepared in an analogous fashion to **1** by Nicholas Wurtz. UV (H_2O) λ_{max} (ε), 312 nm, (66, 600); MALDI-TOF-MS (monoisotopic) [M+H] 1381.75 (calculated 1380.64 for $C_{62}H_{80}N_{26}O_{12}$).

ImPy-β-ImIm-(R)$^{Glu\text{-}NHS}$γ-PyPy-β-ImPy-β-Dp (3-DSG). **3-DSG** was prepared in an analogous fashion to **1-DSG** from **3**. UV (H_2O) λ_{max} (ε), 312 nm, (66, 600); MALDI-TOF-MS (monoisotopic) [M+H] 1592.75 (calculated 1591.69 for $C_{71}H_{89}N_{27}O_{17}$).

ImPy-β-ImIm-(R)$^{(S)\text{-}CBI}$γ-PyPy-β-ImPy-β-Dp (3S). **3S** was prepared in an analogous fashion to **1S** from **3-DSG**. UV (H_2O) λ_{max} (ε), 312 nm, (73, 854); MALDI-TOF-MS (monoisotopic) [M+H] 1781.94 (calculated 1780.76 for $C_{83}H_{101}ClN_{28}O_{16}$).

ImPy-β-ImPy-(R)^{H_2N}γ-PyPy-β-ImPy-β-Dp (4). **4** was prepared in an analogous fashion to **1**. UV (H_2O) λ_{max} (ε), 312 nm, (66, 600); MALDI-TOF-MS (monoisotopic) [M+H] 1380.83 (calculated 1379.65 for $C_{63}H_{81}N_{25}O_{12}$).

ImPy-β-ImPy-(R)^{H_2N}γ-ImPy-β-ImPy-β-Dp (5). **5** was prepared in an analogous fashion to **1**. UV (H_2O) λ_{max} (ε), 312 nm, (66, 600); MALDI-TOF-MS (monoisotopic) [M+H] 1381.9 (calculated 1380.64 for $C_{62}H_{80}N_{26}O_{12}$).

ImPy-β-ImPy-(R)$^{Glu\text{-}NHS}$γ-ImPy-β-ImPy-β-Dp (5-DSG). **5-DSG** was prepared in an analogous fashion to **1-DSG** from **5**. UV (H_2O) λ_{max} (ε), 312 nm, (66, 600).

ImPy-β-ImPy-(R)^{H_2N}γ-ImPy-β-ImPy-β-Dp (5S). **5S** was prepared in an analogous fashion to **1S** from **5-DSG**. UV (H_2O) λ_{max} (ε), 312 nm, (73, 854); MALDI-TOF-MS (monoisotopic) [M+H] 1781.9 (calculated 1780.76 for $C_{83}H_{101}ClN_{28}O_{16}$).

ImIm-β-ImIm-(R)^{H_2N}γ-PyPy-β-PyPy-β-Dp (6). **6** was prepared in an analogous fashion to **1**. UV (H_2O) λ_{max} (ε), 312 nm, (66, 600); MALDI-TOF-MS (monoisotopic) [M+H] 1381.76 (calculated 1380.64 for $C_{62}H_{80}N_{26}O_{12}$).

ImIm-β-ImIm-(R)$^{Glu\text{-}NHS}$γ-PyPy-β-PyPy-β-Dp (6-DSG). 6-DSG was prepared in an analogous fashion to **1-DSG** from **6**. UV (H_2O) λ_{max} (ε), 312nm, (66, 600).

ImIm-β-ImIm-(R)$^{(S)\text{-}CBI}$γ-PyPy-β-PyPy-β-Dp (2R). 2R was prepared in an analogous fashion to **1S** from **6-DSG**. UV (H_2O) λ_{max} (ε), 312nm, (73, 854); MALDI-TOF-MS (monoisotopic) [M+H] 1781.73 (calculated 1780.76 for $C_{83}H_{101}ClN_{28}O_{16}$).

DNA Reagents and Materials. Enzymes were purchased from Boehringer-Mannheim and used with their supplied buffers. Deoxyadenosine 5'-[γ-^{32}P] triphosphates were purchased from I.C.N. Sonicated, deproteinized calf thymus DNA was acquired from Pharmacia. RNase free water was obtained from USB and used for all reactions. All other reagents and materials were used as received. All DNA manipulations were performed according to standard protocols.

Construction of Plasmid DNA. The cDNA plasmids of COX-2 were provided by the Herschman group of UCLA. The HIV-LTR plasmid was provided by the Gottesfeld group at Scripps. The particular plasmid used was prepared by Nicholas Wurtz and contains two mutations that differ from the original plasmid (See notebook 9, page 141 for sequence). Concentration of the prepared plasmid was determined at 260 nm from the relationship of 1 OD unit=50 μg mL^{-1} duplex DNA.

***PCR Labeling* to generate 5'-End-Labeled Restriction Fragments.** To label COX2-198-496 fragment, primers 5'-ATGTCAAAACCGTGGGGAATG-3' and 5'-AGCTTTTGTAACCATAGTGCA-3' were used. For COX-2 fragment 500-772, primers 5'-AAGCCTTCTCCAACCTCTCCT-3' and 5'-TTAAGTCCACTCCATGGCCCA-3' were used. To 5' label the HIV-LTR promoter, primers 5'-AATTCGAGCTCGGTACCC GGT-3' (labeled) and 5'-TGCATGCCTGCAGGTCGATTG-3' (unlabeled) were used. Depending on the strand to be labeled, one primer was treated with T4 polynucleotide kinase and deoxyadenosine 5'-[γ-^{32}P] triphosphate as previously described. PCR reactions containing 60 pmol each primer, 10 µl PCR buffer (Boehringer-Mannheim), 3.7 µl template (0.003 µg/mL), 2 µl dNTP mix (each at 10 mM), 1 µl 100X BSA (New England Biolabs) and 83 µl water were heated at 70 °C for 5 minutes. Four units of Taq Polymerase were added (Boehringer-Mannheim). Thirty amplification cycles were performed, each cycle consisting of the following segments: 94 °C for 1 minute, 54 °C for 1 minute, and 72 °C for 1.5 minutes. Following the last cycle, 10 minutes of extension at 72 °C completed the reaction. The PCR products were gel purified as previously reported for 3'-end labeling protocols.

Preparation of 3'-End labeled Restriction Fragments. The HIV-LTR plasmid was linearized with *EcoRI* and then treated with Sequenase enzyme, deoxyadenosine 5'-α-^{32}P-triphosphate, and thymidine triphosphate for 3' labeling. Treatment with *HindIII* generated the desired 241 base pair fragment which was isolated as previously described.

Cleavage Reactions. All reactions were carried out in a volume of 50 µl. A polyamide stock solution or water (for reference lane) was added to an assay buffer of TE (pH7.5) and 20 kcpm of 3'- or 5'-radiolabeled DNA. The solutions were allowed to equilibrate

for *12 hours* or the appropriate time (for time course reactions) at 37 °C. The reactions were stopped with 60 µl of a solution containing NaOAc (600 mM), EDTA pH 8.0 (12.5 mM), calf thymus DNA (150 µM base pair), glycogen (0.8 mg mL^{-1}), and NaCl (2 M). Ethanol was added to remove unbound polyamide and precipitate the products. The reactions were resuspended in 20 µl of TE (pH 7.5) and cleavage was initiated by heating at 95 °C for 30 minutes. The cleavage products were precipitated with 150 µl ethanol and then resuspended in 100 mM trisborate-EDTA/80% formamide loading buffer, and denatured and loaded onto polyacrylamide gels as previously reported. The gels were quantitated by the use of storage phosphor technology. Yield or efficiency of alkylation was determined as the ratio between the volume integration assigned to the products and the sum of the volumes of all the products in the lane.

Acknowledgements

We are grateful to the National Institutes of Health for research support and the Ralph M. Parsons foundations for a predoctoral fellowship to A.Y.C. We thank Mr. Nicholas Wurtz of the Dervan group for his collaborative efforts on this project. We thank G.M. Hathaway and the Caltech Protein/Peptide Microanalytical Laboratory for MALDI-TOF mass spectrometry.

References

1. Kurumbail, R. G. et al. Structural basis for selective inhibition of cyclooxygenase-2 by anti-inflammatory agents. *Nature* **384**, 644-648 (1996).
2. Garavito, R. M. The cyclooxygenase-2 structure: New drugs for an old target? *Nat. Struct. Biol.* **3**, 897-901 (1996).
3. Kujubu, D. A., Fletcher, B. S., Varnum, B. C., Lim, R. W. & Herschman, H. R. Tis10, a Phorbol Ester Tumor Promoter-Inducible Messenger-Rna from Swiss 3t3 Cells, Encodes a Novel Prostaglandin Synthase Cyclooxygenase Homolog. *J. Biol. Chem.* **266**, 12866-12872 (1991).
4. Wurtz, N. R. & Dervan, P. B. Sequence Specific Alkylation of DNA by Pyrrole-Imidazole Hairpin Polyamide Conjugates. *Chem. Biol.* (1999).
5. Baird, E. E. & Dervan, P. B. Solid phase synthesis of polyamides containing imidazole and pyrrole amino acids. *J. Am. Chem. Soc.* **118**, 6141-6146 (1996).
6. Herman, D. M., Baird, E. E. & Dervan, P. B. Tandem hairpin motif for recognition in the minor groove of DNA by pyrrole - Imidazole polyamides. *Chem.-Eur. J.* **5**, 975-983 (1999).
7. Turner, J. M., Swalley, S. E., Baird, E. E. & Dervan, P. B. Aliphatic/aromatic amino acid pairings for polyamide recognition in the minor groove of DNA. *J. Am. Chem. Soc.* **120**, 6219-6226 (1998).
8. Chang, A. Y. & Dervan, P. B. Strand selective cleavage of DNA by diastereomers of hairpin polyamide-seco-CBI conjugates. *J. Am. Chem. Soc.* **122**, 4856-4864 (2000).

9. Davis, N. A., Majee, S. S. & Kahn, J. D. TATA box DNA deformation with and without the TATA box-binding protein. *J. Mol. Biol.* **291**, 249-265 (1999).

10. Dickinson, L. A. et al. Inhibition of RNA polymerase II transcription in human cells by synthetic DNA-binding ligands. *Proc. Natl. Acad. Sci. USA* **95**, 12890-12895 (1998).

11. Baraldi, P. G. et al. Heterocyclic analogs of DNA minor groove alkylating agents. *Curr. Pharm. Design* **4**, 249-276 (1998).

12. Arcamone, F. M. et al. Synthesis, DNA-Binding Properties, and Antitumor-Activity of Novel Distamycin Derivatives. *J. Med. Chem.* **32**, 774-778 (1989).

13. Baliga, R. et al. Kinetic consequences of covalent linkage of DNA binding polyamides. *Biochemistry* **40**, 3-8 (2001).