Rules for Strand Invasion by Chemically Modified Oligonucleotides

Tsutomu Ishihara and David R. Corey*

Contribution from the Howard Hughes Medical Institute, Departments of Pharmacology and Biochemistry, University of Texas Southwestern Medical Center, 5323 Harry Hines Boulevard, Dallas, Texas 75235-9041

Received November 3, 1998

Abstract: Recognition of sequences within duplex DNA is a general strategy for probing DNA function and for disrupting gene expression. Peptide nucleic acids (PNAs) and conjugates between DNA oligonucleotides and cationic peptides possess superior potential for strand invasion at complementary sequences. To elucidate the rules underlying this phenomenon we examined hybridization to sequences throughout plasmid pUC19. We discovered that oligonucleotide−peptide conjugates and PNAs fall into three classes based on their hybridization efficiencies: (i) those complementary to inverted repeats within AT-rich region hybridize with highest efficiency; (ii) those complementary to areas adjacent to inverted repeats or near AT-rich regions hybridize with moderate efficiency; and (iii) those complementary to other regions do not detectably hybridize, with the exception of PNAs that have been modified to incorporate additional positive charge. Hybridization of oligonucleotide−peptide conjugates and PNAs was stringently dependent on target sequence and was most efficient at sequences within the promoter for β-lactamase or prior to the *Escherichia coli* origin of replication, suggesting that the sequences that regulate biological function may also be among the most susceptible to strand invasion. The correlations between oligomer chemistry, DNA target sequence, and hybridization efficiency that we report here have important implications for the recognition of duplex DNA in cell-free systems and for the selection of target sites for regulating gene expression within cells using synthetic molecules.

Introduction

The development of molecules capable of sequence-specific binding to sites within double-stranded DNA would afford powerful strategies for the control of gene expression that complement approaches that aim to manipulate protein−protein interactions.[1] For single-stranded DNA complementary recognition is usually a straightforward process, easily accomplished using synthetic DNA oligonucleotides. Recognition of duplex DNA by oligonucleotides, however, is more complicated. Hybridization must occur despite pre-existing base-pairing and then be maintained in preference to reannealing of the parent duplex.

The obstacles that confront accessing the information embedded in duplex DNA have led investigators to develop a variety of approaches that permit sequence-specific recognition but avoid disruption of the duplex. These approaches exploit the potential for interactions within the major and minor grooves and include triple helix formation by oligonucleotides,[2] binding of native and modified proteins,[3] and pyrrole−imidazole polyamides.[4] Each strategy has afforded molecules capable of efficient and selective recognition of sequences within chromosomal DNA,[5] demonstrating their potential for accessing information at the genome level. The inherent versatility of Watson−Crick base pairing, however, would provide a valuable complementary method for duplex recognition if two criteria could be met: (i) rules governing access to information with double-stranded DNA must be established, and (ii) molecules capable of sequence-selective binding must be designed to effectively exploit such access. This recognition is termed strand invasion and involves an oligonucleotide or an oligonucleotide mimic binding to its complementary sequence within duplex DNA by Watson−Crick base-pairing, creating a three-stranded complex in which one of the strands of the target duplex is displaced (Figure 1a).

One approach for facilitating hybridization to sequences within duplex DNA is the addition of RecA or similar proteins to promote strand invasion.[6] Another is the use of formamide to destabilize the parent duplex followed by addition of RNA oligomers to create an R-loop.[7] These approaches are effective in cell-free systems, but methods that permit spontaneous strand invasion of supercoiled DNA by single-stranded DNA in the absence of added reagents or protein would facilitate the

* To whom correspondence should be addressed. Telephone: 214-648-5096. Fax: 214-648-5095. E-mail: corey@howie.swmed.edu.

(1) Denison, C.; Kodadek, T. *Chem. Biol.* **1998**, *5*, 129−45.

(2) (a) Moser, H. E.; Dervan P. B. *Science* **1987**, *238*, 645−650. (b) LeDoan, T.; Perrouault, L.; Habhoub, N.; Decout, J.-L.; Thuong, N. T.; Lhomme, J.; Hélène. C. *Nucleic Acids Res.* **1987**, *15*, 7749−7760. (c) Cooney, M.; Czernuszewicz, G.; Postel, E. H.; Flint, S. J.; Hogan, M. E. *Science* **1988**, *241*, 456−459.

(3) (a) Chen, C. B.; Sigman, D. S. *Science* **1987**, *237*, 1197−1201. (b) Sluka, J. P.; Horvath, S. J.; Bruist, M. F.; Simon, M. I.; Dervan, P. B. *Science* **1987**, *238*, 1129−1132. (c) Pei, D.; Schultz, P. G. *J. Am. Chem. Soc.* **1991**, *113*, 9398−9400.

(4) (a) Geierstanger, B. H.; Mrksich, M.; Dervan, P. B.; Wemmer, D. E. *Science* **1994**, *266*, 646−650. (b) Cho, J.; Parks, M. E.; Dervan, P. B. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 10389−10392. (c) Trauger, J. W.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1998**, *120*, 3534−3535. (d) Gottesfeld, J. M.; Neely, L.; Trauger, J. W.; Baird, E. E.; Dervan P. B. *Nature* **1998**, *387*, 202−205.

(5) (a) Strobel, S. A.; Dervan, P. B. *Science* **1990**, *249*, 73−75. (b) Pendergrast, P. S.; Ebright, Y. W.; Ebright, R. H. *Science* **1994**, *265*, 959−962. (c) Gottesfeld, J. M.; Neely, L.; Trauger, J. W.; Baird, E. E.; Dervan, P. B. *Nature* **1997**, *387*, 202−205.

(6) (a) Norirot, P.; Kolodner, R. D. *J. Biol. Chem.* **1998**, *273*, 12274−12280. (b) Cheng, S.; Van Houten, B.; Gamper, H. B.; Sancar, A.; Hearst, J. E. *J. Biol. Chem.* **1998**, *263*, 15110−15117.

(7) (a) Chen, C.-H. B.; Gorin, M. B.; Sigman, D. S. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 4206−4210. (b) Chen, C. H.; Landgraf, R.; Walts, A. D.; Chan, L.; Schlonk, P. M.; Terwilliger, T. C.; Sigman, D. S. *Chem. Biol.* **1998**, *5*, 283−292. (c) Lee, D. Y.; Clayton, D. A. *J. Biol. Chem.* **1996**, *271*, 24262−24269.

10.1021/ja983834e CCC: $18.00
Published on Web 02/24/1999

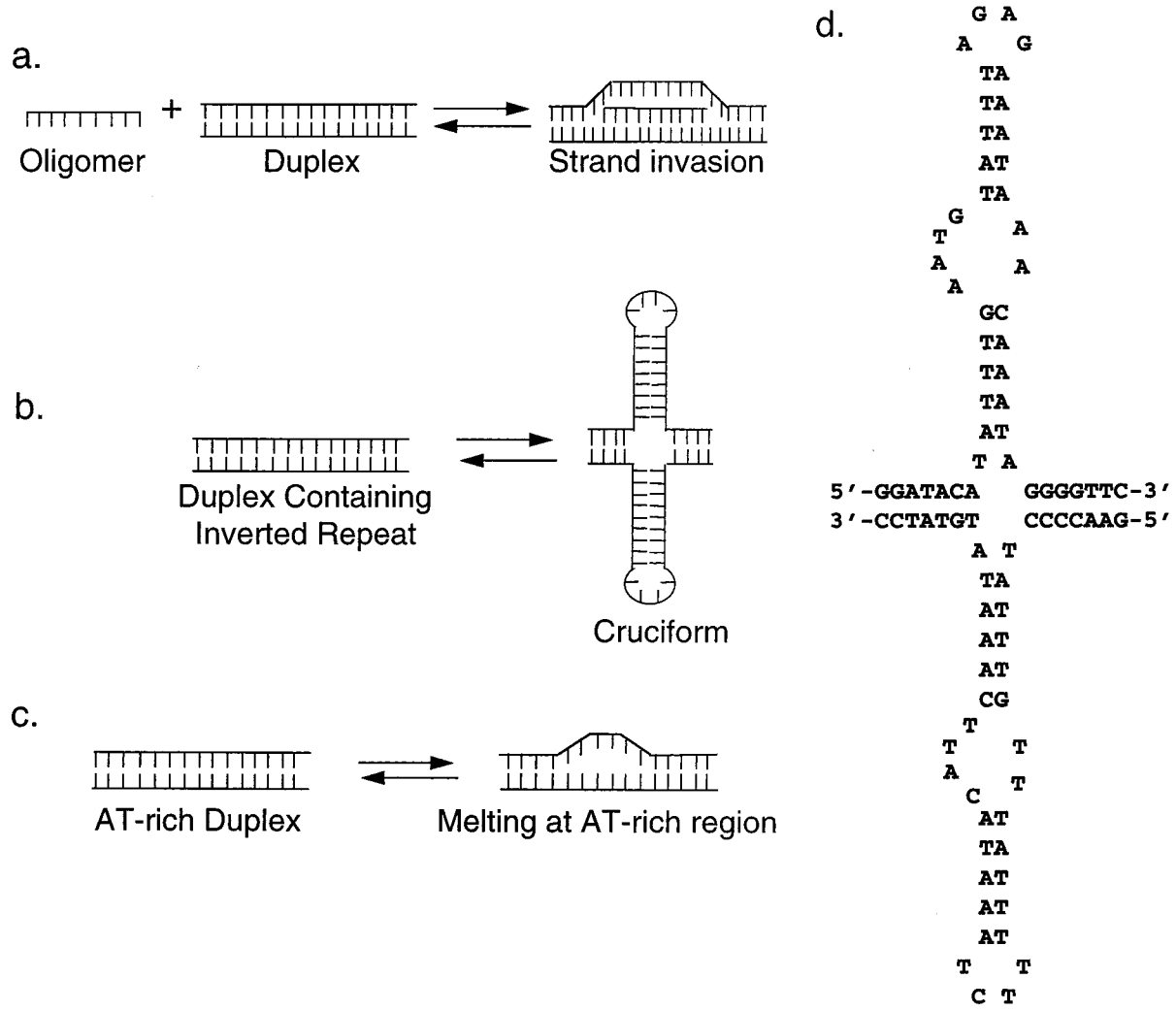

Figure 1. (a) Hybridization of an oligomer to duplex DNA by strand invasion. (b) Formation of a cruciform at at duplex sequence containing an inverted repeat. (c) Melting of DNA at an AT-rich region. Formation of a cruciform or melting of DNA at an AT-rich region leads to unpaired bases that can act to promote strand invasion. (d) Potential cruciform structure formed by the inverted repeat within region 2542−2587 of plasmid pUC19.

development of protocols for targeting sequences within cells and the development of general rules for duplex recognition.

Such spontaneous strand invasion was first recognized by the classic work of Radding and co-workers who noted that supercoiled DNA could spontaneously incorporate complementary single strands.[8] Later Vlassov showed that oligonucleotides could recognize sequences within DNA upon introduction of additional supercoiling in vitro[9] while oligonucleotide−nuclease conjugates can recognize sequences even in the absence of unusually high levels of supercoiling[10] and oligonucleotides have been shown to bind to plasmids at sites expected to form H-DNA structures.[11] Most recently, Gamper and co-workers demonstrated that strand invasion by oligonucleotides can be facilitated by the attachment of triplex-forming guide sequences.[12]

In previous reports we have shown that DNA oligonucleotides can hybridize to inverted repeat sequences within normally supercoiled DNA.[13] This hybridization can be promoted by attachment of a cationic protein[13,14] or cationic peptides[15] to the oligonucleotide, with the rate association constant k_a for conjugate binding being increased up to 48 000-fold for oligonucleotide−peptide conjugates relative to unmodified oligonucleotides. This enhanced hybridization was conferred by the reduction in overall negative charge by the attached peptide, leading us to examine hybridization by peptide nucleic acids (PNAs) (Figure 2)[16] in which overall negative charge is reduced by the use of unchanged backbone linkages.

PNAs possess a nonionic backbone in which the deoxyribose linkages have been replaced by *N*-(2-aminoethyl)glycine units.[16] The uncharged nature of the PNA internucleotide linkages increases their affinity for complementary sequences under

(8) (a) Holloman, W. K.; Wiegand, R.; Hoessli, C.; Radding, C. M. *Proc. Natl. Acad. Sci. U.S.A.* **1975**, *72*, 2394−2398. (b) Beattie, K. L.; Wiegand, R. C.; Radding, C. M. *J. Mol. Biol.* **1977**, *116*, 783−803.

(9) Vlassov, V. V.; Gaidamakov, S. A.; Zarytova, V. F.; Knorre, D. G.; Levina, A. S.; Nikonova, A. A.; Podust, L. M.; Fedorova, O. S. *Gene* **1988**, *72*, 313−322.

(10) Corey, D. R.; Pei, D.; Schultz, P. G. *J. Am. Chem. Soc.* **1989**, *111*, 8523−8525.

(11) Belotserkovskii, B. P.; Krasilnikova, M. M.; Veselkov, A. G.; Frank-Kameneteskii, M. D. *Nucleic Acids Res.* **1992**, *20*, 1903−1908.

(12) Gamper, H. B.; Hou, Y.-M.; Stamm, M. R.; Podyminogin, M. A.; Meyer, R. B. *J. Am. Chem. Soc.* **1998**, *120*, 2182−2183.

(13) Iyer, M.; Norton, J. C.; Corey, D. R. *J. Biol. Chem.* **1995**, *270*, 14712−14717.

(14) Corey, D. R.; Munoz-Medellin, D.; Huang, A. *Bioconjugate Chem.* **1995**, *6*, 93−100.

(15) Corey, D. R. *J. Am. Chem. Soc.* **1995**, *117*, 9373−9374.

(16) (a) Nielsen, P. E.; Egholm, M.; Berg, R. F.; Buchardt, O. *Science* **1991**, *254*, 1497−1500. (b) Hanvey, J. C.; Peffer, N. J.; Bisi, J. E.; Thomson, S. A.; Cadilla, R.; Josey, J. A.; Ricca, D. J.; Hassman, C. F.; Bonham, M. A.; Au, K. G.; Carter, S. G.; Bruckenstein, D. A.; Boyd, A. L.; Noble, S. A.; Babiss, L. E. *Science* **1992**, *258*, 1481−1485. (c) Nielsen, P. E. *Biophys. Chem.* **1997**, *68*, 103−108.

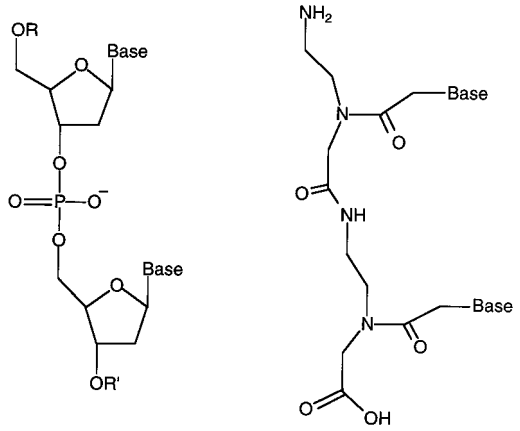

Figure 2. Chemical structures of DNA and PNA oligomers.

conditions of low ionic strength[17] and increases the rate of their hybridization,[18] leading to their use in a growing number of applications.[19] PNAs can hybridize to homopurine−homopyrimidine sequences within duplex DNA by a four-stranded P-loop complex composed of a PNA−PNA−DNA triplex and a displaced DNA strand.[20]

Rapid, high affinity hybridization and the ability to form P-loops make PNAs excellent candidates for recognition of duplex DNA, although the sequence-dependence for their hybridization or that of oligonucleotide−peptide conjugates remains unclear. RNA and genomic DNA within cells contain important targets for recognition, and recently several methods for efficient delivery of PNAs within eukaryotic[21−24] and prokaryotic[25] cells have been developed. These include attachment to import peptides, transient permeabilization of membranes using streptolysin O,[22] electroporation,[23] and delivery of DNA/PNA hybrids.[24]

Here we explore the relative susceptibilities of sequences throughout a supercoiled plasmid for hybridization by strand invasion. We scanned plasmid pUC19 using oligonucleotide−peptide conjugates to identify regions that are the most permissive for strand invasion under moderate conditions. We then evaluated hybridization of PNAs to varied sequences and used it to either disrupt or promote hybridization of oligonucleotide−peptide conjugates directed either to adjacent or to distant regions of plasmid. Sequences that are likely to be at least partially single-stranded are present throughout DNA,[26−29] particularly promoter regions,[26−28] and triplet repeats,[29] and our observations will help develop the rules for a knowledge-based approach aimed at gaining efficient access to them. The advent of methods for intracellular PNA delivery enhance the likelihood that these rules will not only guide recognition in cell-free systems but can also be applied to the targeting of sequences within cells.

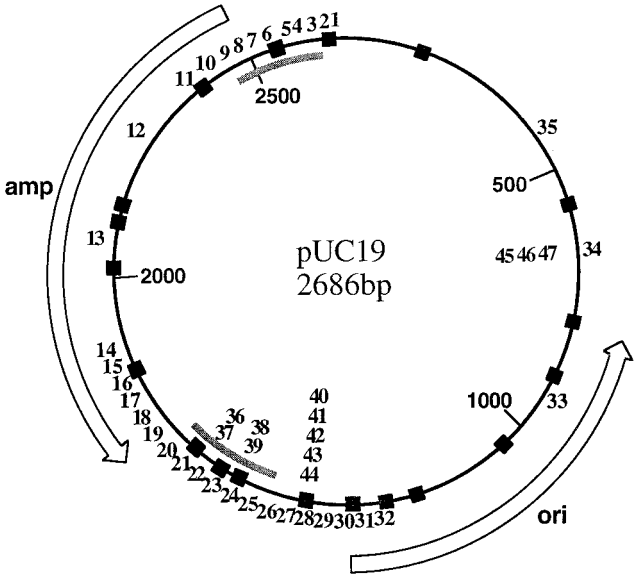

Figure 3. Location of target sites for hybridization of DNA oligonucleotide−peptide conjugates, PNAs, and PNA derivatives within pUC19. Inverted repeats contain three to ten nucleotides in the loop regions and at least five nucleotides in the stem. Black boxes denote approximate location of sequences with inverted repeats. Gray lines show the AT-rich regions (over 60% AT base-pairs relative to GC basepairs) within pUC19. The location of the ampicillin gene and the *E. coli* origin of replication are noted.

Results

Scanning pUC19 for Hybridization by DNA Oligonucleotide−Peptide Conjugates. To establish rules for strand invasion we synthesized 35 disulfide-linked conjugates between DNA oligonucleotides complementary to sequences throughout supercoiled pUC19[30] (Figure 3) (Table 1) and the cationic peptide CAAKKAAKKAAKKAAKK. Conjugates were chosen to span sites that (i) contained inverted repeats, (ii) were within regions characterized by a high proportion of adenosine−thymidine base-pairs (i.e., AT-rich), or (iii) had no clear propensity to adopt

(17) (a) Egholm, M.; Buchardt. O.; Christensen. L.; Behrens. C.; Freier. S. M.; Driver, D. A.; Berg, R. H.; Kim, S. K.; Norden, B.; Nielsen, P. E. *Nature* **1993**, *365*, 566−568. (b) Hamilton, S. E.; Pitts, A. E.; Katipally, R. R.; Jia, X.; Davies, B. A.; Rutter, J. P.; Wright, W. R.; Shay, J. W.; Corey, D. R. *Biochemistry* **1997**, *36*, 11873−11880.

(18) Smulevitch, S. V.; Simmons, C. G.; Norton, J. C.; Wise, T. W.; Corey, D. R. *Nature Biotech.* **1996**, *14*, 1700−1704.

(19) Corey, D. R. TIBTECH **1997**, *15*, 224−229.

(20) (a) Lomakin, A.; Frank-Kamenetskii, M. D. *J. Mol. Biol.* **1998**, *276*, 57−70. (b) Demidov, V. V.; Yavnilovich, M. V.; Belotserkovskii, B. P.; Frank-Kamenetskii, M. D.; Nielsen, P. E. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 2637−2641. (c) Demidov, V. V.; Yanilovich, M. V.; Frank-Kamenetskii, M. D. *Biophys. J.* **1997**, *72*, 2763−2769. (d) Nielsen, P. E.; Christensen, L. *J. Am. Chem. Soc.* **1996**, *118*, 2287−2288.

(21) (a) Simmons, C. G.; Pitts, A. E.; Mayfield, L. D.; Shay, J. W.; Corey, D. R. *Bioorg. Med. Chem. Lett.* **1997**, *7*, 3001−3007. (b). Basu, S.; Wickstrom, E. *Bioconjugate Chem.* **1997**, *8*, 481−488. (c) Pooga, M.; Ursel, S.; Hallbrink, M.; Valkna, A.; Saar, K.; Rezaei, K.; Kahl, U.; Hao, J.-K.; Xu, X.-J.; Wiesenfeld-Hallin, Z.; Hokfelt, T.; Bartfai, T.; Langel, U. *Nature Biotech.* **1998**, *16*, 857−861.

(22) Faruqui, A. F.; Egholm, M.; Glazer, P. M. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *95*, 1398−1403.

(23) Shammas, M. A.; Simmons, C. G.; Corey, D. R.; Shmookler Reis, R. J., unpublished results.

(24) Hamilton S. E.; Corey D. R., unpublished results.

(25) (a) Good, L.; Nielsen, P. E. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 2073−2076. (b) Good, L.; Nielsen, P. E. *Nature Biotech.* **1998**, *16*, 355−358.

(26) (a) Pearson C. E.; Zorbas, H.; Price, G. B.; Zannis-Hadjopoulos, M. J. *Cell. Biochem.* **1996**, *63*, 1−22. (b) Schroth, G. P.; Ho, S. *Nucleic Acids Res.* **1995**, *23*, 1977−1983. (c) Marilley, M.; Pasero, P. *Nucleic Acids Res.* **1996**, *24*, 2204−2211. (d) Kowalski, D.; Natale, D. A.; Eddy, M. J. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *85*, 9464−9468. (e) Potaman, V. N.; Ussery, D. W.; Sinden, R. R. *J. Biol. Chem.* **1996**, *271*, 13441−13447. (f) Kim, E. L.; Peng, H.; Esparza, F. M.; Maltchenko, S. Z.; Stachowiak, M. K. *Nucleic Acids Res.* **1998**, *26*, 1793−1800. (g) Dai, X.; Greizerstein, M. B.; Nadas-Chinni, K.; Rothman-Denes, L. B. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 2174−2179. (h) Tismit, Y.; Moras, D. *Q. Rev. Biophys.* **1996**, *29*, 279−307.

(27) (a) Benham, C. J. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 2999−3003. (b) Benham C. J. *J. Mol. Biol.* **1996**, *255*, 425−434. (c) Benham, C. *J. Comput. Appl. Biosci.* **1996**, *12*, 375−381.

(28) Dai, X.; Kloster, M.; Rothman-Denes, L. B. *J. Mol. Biol.* **1998**, *283*, 43−58.

(29) (a) Pearson, C. E.; Sinden, R. R. *Curr. Opin. Struct. Biol.* **1998**, *8*, 321−330. (b) Ohshima, K.; Kang, S.; Larson, J. E.; Wells, R. D. *J. Biol. Chem.* **1996**, *271*, 16773−16783. (c) Schumacher, S.; Fuchs, R. P. P.; Bichara, M. *J. Mol. Biol.* **1998**, *279*, 1101−1101.

(30) Yanisch-Perron, C.; Viera, J.; Messing, J. *Gene* **1985**, *33*, 103−119.

Table 1. DNA Oligonucleotide−Peptide Conjugates[a]

	sequence	location within pUC19		sequence	location within pUC19
1	GCGAGACGAAAGGGCCTC	2673−(2686)−4	**19**	ACGAAATAGACAGATCGC	1660−1677
2	TTATAGGTTAATGTCATG	2640−2657	**20**	TAGGTGCCTCACTGATTA	1637−1654
3	TAGACGTCAGGTGGCACT	2607−2624	**21**	TGGTAACTGTCAGACCAA	1614−1631
4	CGGGGAAATGTGCGCGGAA	2585−2603	**22**	CTCATATATACTTTAGAT	1591−1608
5	CGCGGAACCCCTATTTGT	2574−2591	**23**	TTTAAAACTTCATTTTTA	1570−1587
6	GAGCGGATACATATTTGAATG	2538−2558	**24**	GGATCTTCACCTAGATCCT	1544−1562
7	TGATAAATGCTTCAATAA	2505−2522	**25**	GATAATCTCATGACCAAA	1521−1538
8	GAAAAAGGAAGAG	2488−2500	**26**	TTAACGTGAGTTTTCGTT	1498−1515
9	AGTATGAGTATTCAACAT	2472−2489	**27**	GAGCGTCAGACCCCGTAG	1475−1492
10	ATTCCCTTTTTTGCGGCA	2439−2456	**28**	ATCAAAGGATCTTCTTGA	1452−1469
11	TTTGCTCACCCAGAAACG	2406−2423	**29**	CTGCGCGTAATCTGCTGC	1422−1439
12	GCCACATAGCAGAACTTTAA	2259−2278	**30**	AACAAAAAAACCACCGCT	1399−1416
13	GTAAGTTGGCCGCAGTGTTA	2081−2100	**31**	CGGTGGTTTGTTTGCCGG	1376−1393
14	CGGATAAAGTTGCAGGAC	1838−1855	**32**	GAGCTACCAACTCTTTTT	1353−1370
15	CGGCCCTTCCGGCTGGCT	1808−1825	**33**	CAGGGTCGGAACAGGAGA	988−1005
16	ATAAATCTGGAGCCGGTG	1778−1795	**34**	ACCGCCTCTCCCCGCGCG	652−669
17	GCGGTATCATTGCAGCAC	1748−1765	**35**	CATGCCTGCAGGTCGACT	428−445
18	ACGATACGGGAGGGCTTACC	1717−1736			

[a] All oligonucleotide−peptide conjugates (**1**−**35**) are listed 5′ to 3′ termini. Peptide (LysLysAlaAlaLysLysAlaAlaLysLysAlaAlaLysLysAlaAlaCys) is conjugated at the 5′ termini of each oligonucleotide through the disulfide bond.

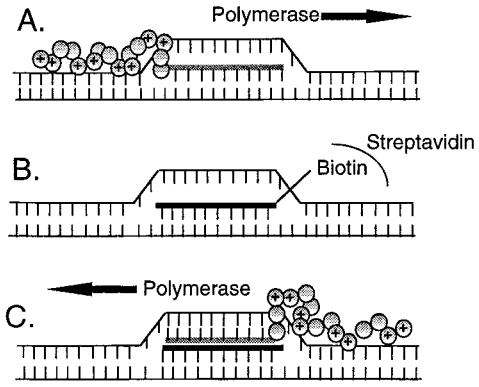

Figure 4. Assays for strand invasion of plasmid DNA by chemically modified oligomers. (A) Hybridization of an oligonucleotide−peptide conjugate. Hybridization is detected by the addition of DNA polymerase and the subsequent use of the oligonucleotide−peptide conjugate as a primer. The efficiency of hybridization is quantitated by phosphorimager analysis, and the location of hybridization is deduced from analysis of the sequence data. (B) Hybridization of a biotin-labeled PNA conjugate. Hybridization is detected by removing the hybridized PNA−plasmid complex with streptavidin-linked magnetic beads. (C) Hybridization of a PNA followed by hybridization of an oligonucleotide−peptide conjugate. Hybridization of the PNA is detected by monitoring the ability of the oliognucleotide−peptide conjugate to hybridize and act as a primer for DNA polymerase. Hybridization of the conjugate is performed under conditions (usually low temperature) that preclude binding in the absence of an open region created by PNA hybridization.

non-B-type secondary structure. DNA sequences that were AT-rich or that contained inverted repeats were chosen for special attention because they are more likely to form secondary structures such as cruciforms (Figure 1b) or bubbles (Figure 1c) that contain unpaired bases that might favor initiation and subsequent maintenence of strand invasion by a complementary oligomer. Supercoiled DNA was used because supercoiling is present in DNA that is transcriptionally active and because negative superhelical tension promotes transient formation of single stranded regions and is known to enhance uptake of polypyrimidine PNAs.[31]

Conjugates that hybridize by strand invasion can act as primers for modified T7 DNA polymerase (Figure 4A), allowing hybridization to be evaluated by monitoring the products of strand elongation by either polyacrylamide or agarose gel electrophoresis.[13,18] Polyacrylamide gel electrophoresis separates the elongation products to base-pair resolution and affords sequence information that unambiguously confirms the location of hybridization. The relative efficiencies of hybridization can be determined by quantification of the radiolabeled products by phosphorimager analysis (Figure 5 a and c). Strand elongation of primer modifies the mobility of the template plasmid DNA during agarose gel electrophoresis (Figure 5b). Comparison of the amount of plasmid with altered and unaltered mobility sets a lower limit on the absolute efficiency of hybridization.

Hybridization at Regions that are AT-rich and/or Contain Inverted Repeats. Conjugates **21**−**24** hybridized to pUC19 (Figure 3), with efficiencies ranging from >70% absolute efficiency for conjugate **24** to 10−15% for conjugate **23**. The region spanned by these conjugates is AT-rich (75%) and contains three inverted repeats (Figures 3 and 6). These inverted repeats span the binding sites of conjugates **21**, **23**, and **24** and have the potential to form five-, six-, and 10-base stem regions respectively (Figure 6). Conjugate **24** was targeted to an inverted repeat at bases 1540−1565. This inverted repeat is the most prominent site for nicking by S1 nuclease, and the high efficiency of strand invasion at this site is probably due to a propensity for it to extrude a cruciform structure in which bases at the base of the stem region and in the loop region are unpaired. Conjugate **22** did not target an inverted repeat, and its hybridization is probably due to disruption of the DNA structure caused by the presence of adjacent inverted repeats as well as by the AT-rich nature (77%, 14/18 bases) of its target site.

Strand elongation by conjugates **18**−**20**, **25**, **26**, and **28** was also detected but was less efficient than that of conjugates **21**−**24** (Figure 5). Less efficient hybridization by conjugates **20** and **25** is striking, because their target sites lie only 10 bases beyond the sites for conjugates **21** and **24** (Figure 6). No distinct strand elongation was detected upon addition of the other conjugates (**14**−**17**, **27**, **29**−**32**). It is interesting to note that the conjugates whose hybridization was detectable bind near the region spanned by conjugates **21**−**24** that is both AT-rich and contains multiple inverted repeats, **18**−**20**, **25**, and **26**, or hybridize to a separate inverted repeat, **28** (Figure 5c, Figure 6). Hybridization by conjugates **18**−**20** suggests that the combination of AT-rich sequence and inverted repeats not only facilitates direct strand invasion but also tend to promote strand invasion at sequences up to 100 bases distant.

(31) Bentin, T.; Nielsen, P. E. *Biochemistry* **1996**, *35*, 8863−8869.

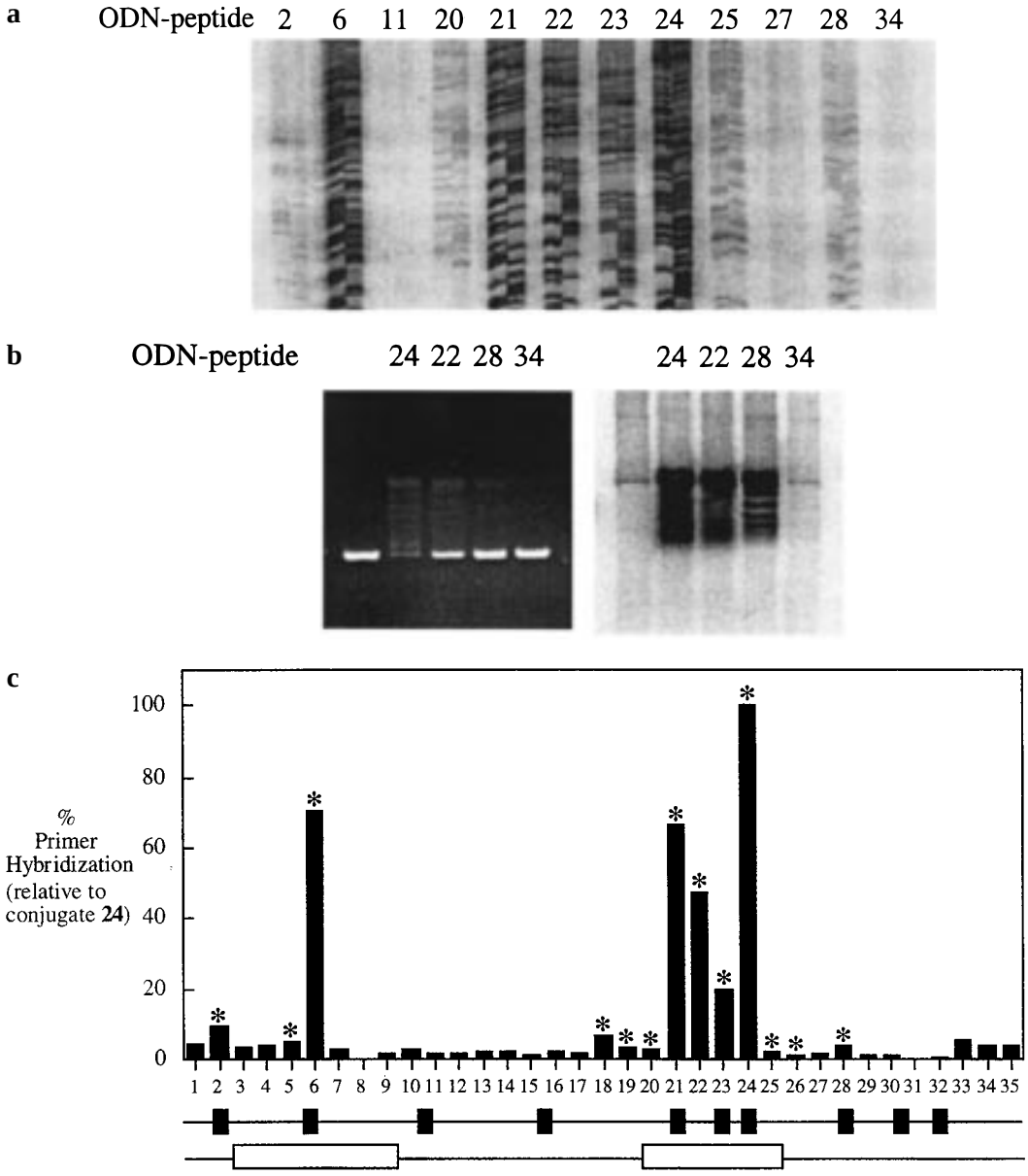

Figure 5. (a) Strand elongation by representative DNA oligonucleotide−peptide conjugates monitored by 6% polyacrylamide gel electrophoresis. pUC19 (40 nM) was mixed with 20 equiv of various oligonucleotide peptide conjugates prior to the addition of modified T7 DNA polymerase and strand elongation. (b) Strand elongation using DNA oligonucleotide−peptide conjugates monitored by 1% agarose gel electrophoresis. The identity of the conjugate added is noted. Altered mobility of pUC 19 was visualized on agarose gel by either ethidium bromide staining or autoradiography. (c) Quantification of strand elongation by DNA oligonucleotide−peptide conjugates by phosphorimager analysis. Values are relative to strand elongation using conjugate **24** as a primer. Black boxes and white boxes denote the approximate location of inverted repeat regions and AT-rich regions (over 60%), respectively. Asterisks mark conjugates that yield elongation products that are detectable above the background.

Bases 2549−2580 within pUC19 have the potential to form an inverted repeat containing an 11-base stem region consisting of five- and six-base stems separated by a short mismatched region (two bases on one strand, four on the other) (Figure 1d). This sequence is also within a region that is relatively AT-rich; therefore, to gain additional insights into the rules governing strand invasion, we designed a series of oligonucleotide−peptide conjugates, **1**−**11** (Figure 3, Table 1), to span this inverted repeat and the surrounding region.

Elongation of conjugates **1**−**11** directed to sequences surrounding the 2549−2580 inverted repeat was more stringently dependent on sequence than was hybridization to the region surrounding 1540−1565 repeat (Table 1). Conjugate **6**, which was targeted to the inverted repeat, hybridized with an efficiency similar to that displayed by conjugates **21**−**24**. Conjugate **5**, which was targeted to a sequence directly adjacent, supported only a low level of strand elongation, as did conjugate **2** which was target to an inverted repeat at bases 2650−2665. Addition of the other conjugates targeted to this region, **1**, **3**, **4**, and **7**−**11**, did not result in distinct strand elongation, even though conjugate **7** was targeted to a sequence only 16 bases distant from the target sequence for conjugate **6**. We also assayed conjugates **12**, **13**, and **33**−**35** that were targeted to regions of pUC19 that did not contain inverted repeats or AT-rich regions and found that they did not act as primers, supporting the suggesting that strand invasion is stringently sequence-dependent.

Hybridization of PNAs to Sequences Within pUC19. To generalize the rules governing strand invasion of duplex DNA to another class of chemically novel oligomers, we examined hybridization by PNAs. We used two complementary assays to monitor binding of PNAs to plasmid. The first exploits the ability of biotin-labeled PNAs to remove bound plasmid from

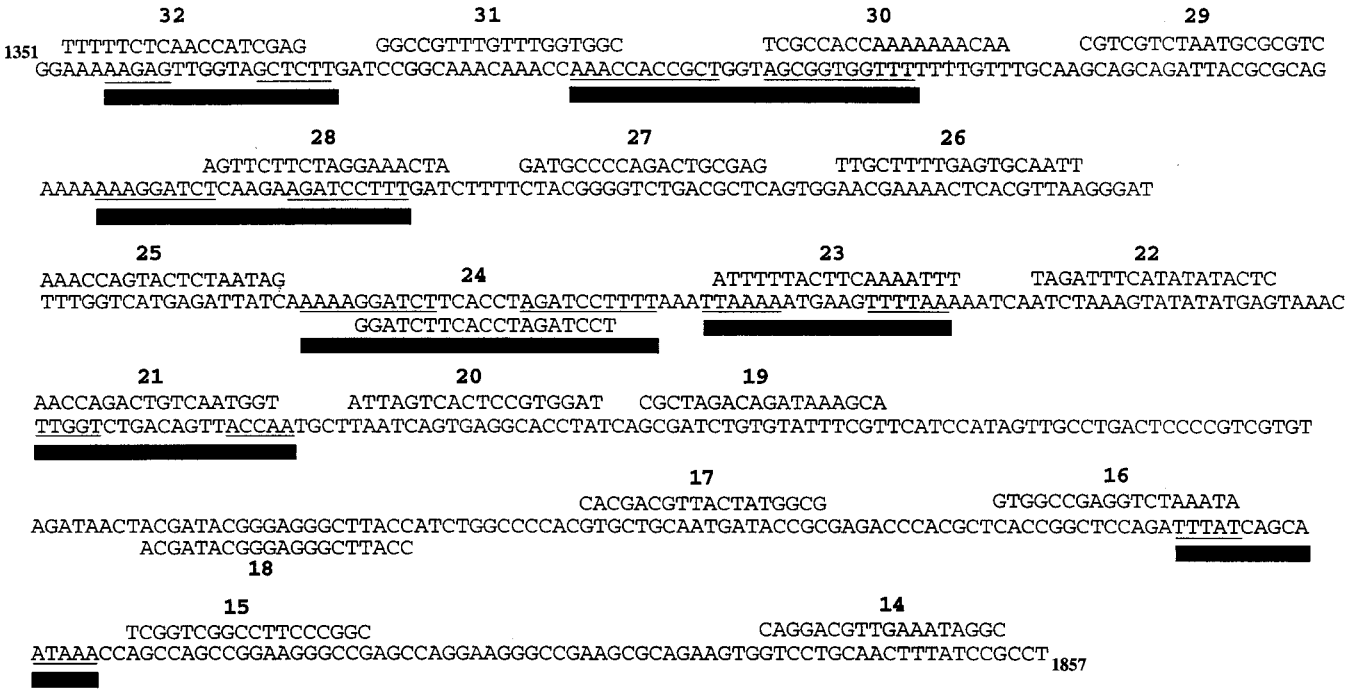

Figure 6. Sites for hybridization of conjugates **14**−**32**. Black boxes denote location of inverted repeats, while underlined sequences denote the inverted repeats themselves.

solution upon treatment with streptavidin beads (Figure 4B). This elegant technique has been used for affinity capture of chromosomal DNA by PNAs directed to polypurine−polypyrimidine sequenes[32] and is a direct test for hybridization. The second assay uses the strand that is displaced by PNA binding as a target for hybridization of an oligonucleotide conjugate (Figure 4C), thereby confirming the sequence specificity of PNA binding and allowing us to evaluate the extent to which PNA binding opens adjacent regions of the duplex for hybridization. While not a direct assay of hybridization, the hybridization of the oligonucleotide−peptide conjugate to the displaced strand is a highly sensitive probe, capable of detecting low levels of strand invasion. Another advantage of our assay is that, as noted above, polymerization of primer yields DNA sequence information that unambiguously confirms the location of both primer and PNA binding.

Affinity Capture of pUC19 by Biotin-Labeled PNAs. Upon incubation with streptavidin, biotin-labeled PNAs **37** and **39** that were analogous to conjugates **22** and **24** allowed plasmid pUC19 to be removed from solution (Figure 7). Capture by **37** was less efficient than capture by **39**, consistent with our previous observation that strand elongation by **22** was less efficient than strand elongation by **24**. No capture was observed upon addition of biotin labeled PNA **48**, which was not complementary to pUC19.

We also tested PNA **43** which was analogous in sequence to conjugate **28** and observed affinity capture but, as would be expected from the low level of strand elongation noted upon addition of **28**, the efficiency was low relative to capture by **39**. To determine whether chemical modifications to PNAs might further enhance strand invasion, a possibility already noted for lysine-containing bis-PNAs that bind by triplex formation,[33] we synthesized PNA **44** which was analogous in sequence to **43** but which is attached to a cationic peptide and observed that this modification substantially enhanced affinity capture

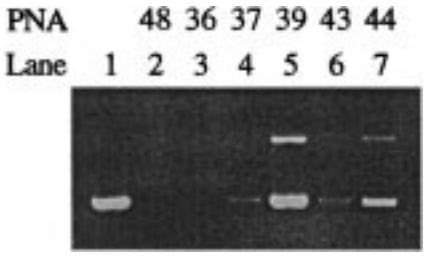

Figure 7. Affinity capture of plasmid DNA by biotin-labeled PNAs. Plasmid DNA bound on Dynabeads using biotin-labeled PNAs was visualized on 1% agarose gel. Lane 1, no PNA or streptavidin-labeled matrix added; lane 2, biotin-labeled PNA **48**, which is not complementary to pUC19, added; lane 3, PNA **36**, which lacks biotin, added; lane 4, biotin-labeled PNA **37** added; lane 5, biotin-labeled PNA **39** added; lane 6, biotin-labeled PNA **43** added, lane 7, biotin-labeled PNA−peptide chimera **44** added.

when compared to **43** (Figure 7). The band of lower mobility visible upon plasmid capture by **39** and **44** is nicked plasmid, presumably formed during the 18 h incubation used to bind the PNA−plasmid complex to the streptavidin coated beads, evidence that PNA hybridization is stable upon relaxation of supercoiling.

Promotion of Hybridization of Oligonucleotide−Peptide Conjugate by PNA Addition. As mentioned above, strand invasion by PNAs creates a displaced strand that can bind to complementary oligonucleotide−peptide conjugates. To map the extent to which this displacement affects accessibility we annealed PNA **40** with pUC19 and then evaluated the promotion of strand elongation by conjugates targeted to adjacent sequences. Annealing was carried out at 65 °C rather than at 37 °C, the standard temperature for all of the other studies described in this work, because this improves the efficiency of hybridization by 3-fold, allowing effects to be more clearly visualized and quantitatively evaluated.

(32) Bukanov, N. O.; Demidov, V. V.; Nielsen, P. E.; Frank-Kamenetski, M. D. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 5516−5520.

(33) (a) Griffith, M. C.; Risen, L. M.; Greig, M. J.; Lesnik, E. A.; Sprankle, K. G.; Griffey, R. H.; Kiely, J. S.; Freier, S. M. *J. Am. Chem. Soc.* **1995**, *117*, 831−832. (b) Kuhn, H.; Demidov, V. V.; Frank-Kamenetskii, M. D.; Nielsen, P. E. *Nucleic Acids Res.* **1998**, *26*, 582−587.

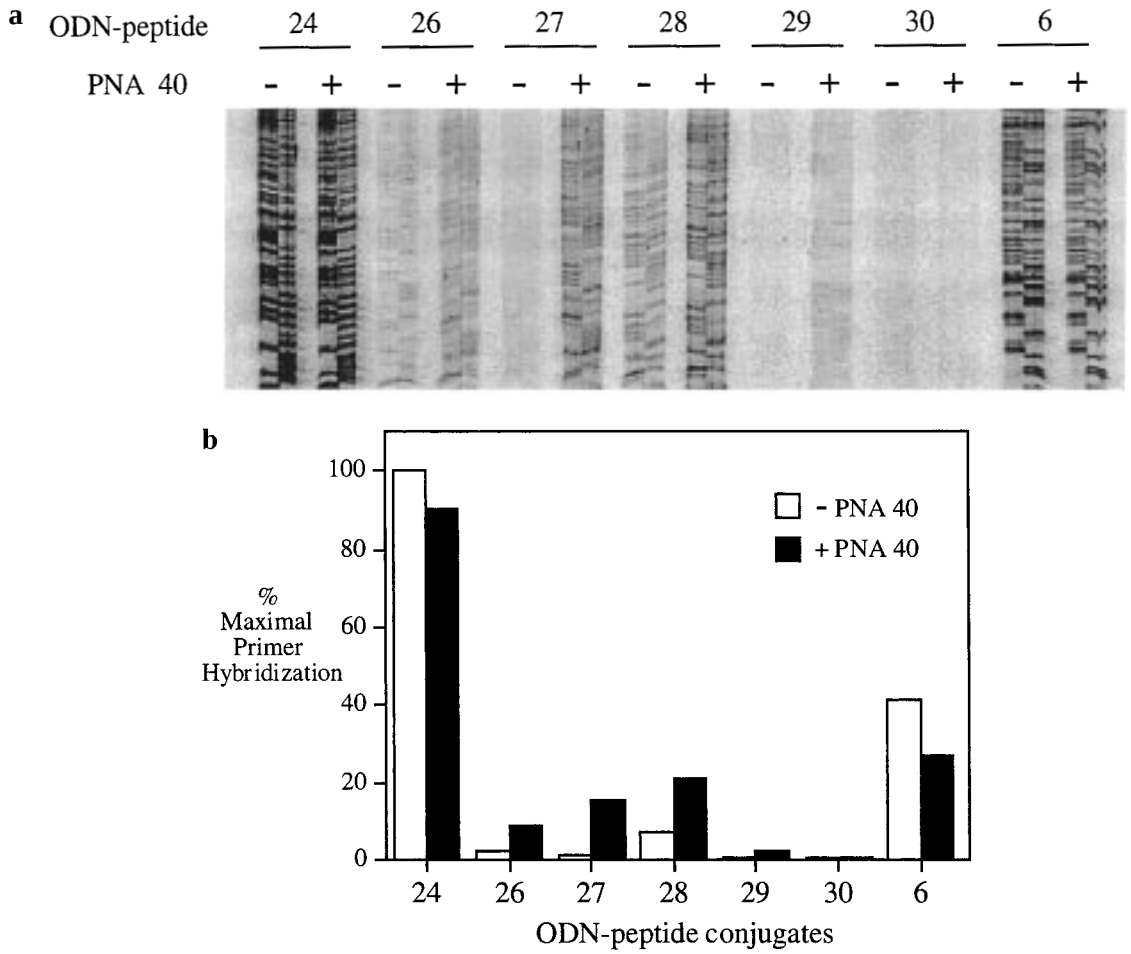

Figure 8. Influence of target sequence on complementary-assisted hybridization using PNA. 40 nM pUC 19 were incubated with or without 200 nM PNA 40 for 5 min at 65 °C and followed by 800 nM various conjugates for 15 min at 37 °C. The elongation products by Sequenase were visualized on 6% polyacrylamide gel (a) and quantified by phosphorimager (b). Values are relative to strand elongation using conjugate **24** without addition of PNA **40**.

Table 2. PNAs and PNA Derivatives[a]

	sequence	location within pUC19	analogues to conjugate
36	Gly-AGGATCTAGGTGAAGATC-Lys	1591−1608	**22**
37	Biotin-AEEA-AEEA-AGGATCTAGGTGAAGATC-Lys	1591−1608	**22**
38	Gly-ATCTAAAGTATATATGAG-Lys	1545−1562	**24**
39	Biotin-AEEA-AEEA-ATCTAAAGTATATATGAG-Lys	1545−1562	**24**
40	Gly-TCAAGAAGATCCTTTGAT-Lys	1452−1469	**28**
41	Gly-AGGATCTCAAGAAGATCCTTTGATCTTTTC-Lys	1446−1475	**28**
42	Peptide-Gly-TCAAGAAGATCCTTTGAT-Lys	1452−1469	**28**
43	Biotin-AEEA-AEEA-TCAAGAAGATCCTTTGAT-Lys	1452−1469	**28**
44	Biotin-AEEA-AEEA-Peptide-Gly-TCAAGAAGATCCTTTGAT-Lys	1452−1469	**28**
45	Gly-CGCGCGGGGAGAGGCGG-Lys	652−668	**34**
46	Gly-GGCCAACGCGCGGGGAGAGGCGGTTTGCGT-Lys	646−675	**34**
47	Peptide-Gly-CGCGCGGGGAGAGGCGG-Lys	652−668	**34**
48	Biotin-AEEA-AEEA-TGCTCTAGAATGAAC-Lys	mismatch	

[a] All PNAs and PNA derivatives (**36**−**48**) are listed N to C termini. AEEA is 2-aminoethoxy-2-ethoxy acetic acid. Peptide is LysLysAlaAla-LysLysAlaAlaLysLysAlaAlaLysLys.

Addition of PNA **40** led to enhanced priming by subsequently added conjugates, not only by **28**, which was fully complementary to PNA **40** and to the strand it displaced, but also by conjugates **26**, **27**, and **29** which were complementary to adjacent sequences (Figure 8 a, b). Hybridization was enhanced further when the cationic peptide−PNA chimera **44** (Table 2) was added (results not shown). These results indicate that the displacement created by PNA binding not only affects the sequence at the binding site but affects the accessibility of nearby sequences as well. Use of PNA **41**, which was 30 bases long, yielded similar results to 18-base PNA **40**, demonstrating that simply lengthening the PNA was not a helpful strategy for enhancing strand invasion at this sequence.

Noting that PNA−peptide chimera possess an enhanced ability to promote strand invasion, we sought to determine whether their hybridization might also allow oligonucleotide−peptide conjugates to prime strand elongation at sequences that did not contain inverted repeats and were not within AT-rich regions. We synthesized PNA−peptide chimera **47** to hybridize at a site that was neither an inverted repeat nor was within an AT-rich region (Table 2). For comparison we also obtained the analogous 18- and 30-nucleotide PNAs, **45** and **46** that lacked

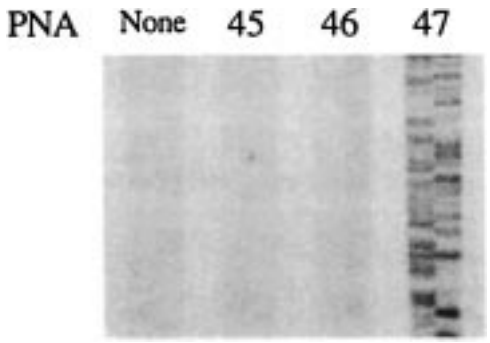

Figure 9. Effect of PNA modifications on complementary-assisted hybridization. 40 nM pUC19 were incubated with 200 nM PNA **45**, **46**, or **47** for 30 min at 37 °C and followed by 800 nM conjugate **34** for 15 min at 37 °C. The elongation products by Sequenase were visualized on 6% polyacrylamide gel.

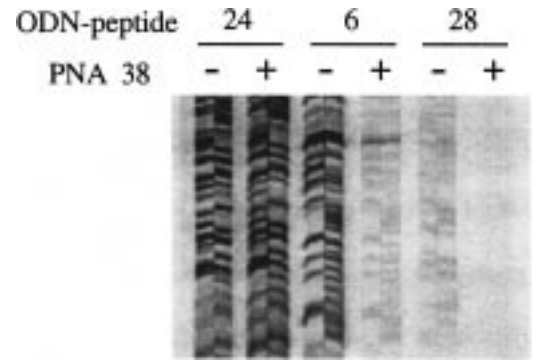

Figure 10. Action at a distance upon PNA hybridization. 40 nM pUC-19 were incubated with or without 200 nM PNA **38** for 5 min at 37 °C and followed by 800 nM conjugates **6**, **24**, or **28** for 15 min at 37 °C. The elongation products produced upon addition of Sequenase were visualized using a 6% polyacrylamide gel. Similar results were obtained upon addition of PNA **36**.

the peptide, and observed that PNA−peptide **47** was able to promote strand elongation (Figure 9) whereas PNA **45** and **46** did not. This result demonstrates that addition of cationic charge can permit strand invasion at sequences that would otherwise have been inaccessible, extends earlier observations on enhanced strand invasion through the addition of a positive charge to PNAs targeted to polypurine−polypyrimidine sequences,[33] and suggests that this improved hybridization upon addition of positive charge will be a general rule for diverse sequences.

We noted that the PNA component of PNA−peptide **47** has a nine-base homopurine stretch and that it is possible that triplex formation may play a role in initiating strand recognition. The fact that we observed hybridization of the complementary oligonucleotide−peptide conjugate, however, indicates that the complex we detected must have involved strand invasion. Hybridization by PNA−peptide **47**, while modest, demonstrates that chemical modification of PNAs can lead to strand invasion at sequences that where hybridization was previously undetectable and suggests that other modifications may exist that will further enhance the efficiency of strand invasion.

Ability of PNA Hybridization to Influence DNA Structure at a Distance. As noted above, hybridization of PNAs can promote hybridization of oligonucleotide conjugates. Conversely, we have previously reported that PNAs are able to block hybridization of oligonucleotide−peptide conjugates with the same sequence.[13,18] We now sought to determine whether PNAs might be able to exert an effect at a distance and block hybridization to sites hundreds or thousands of bases away from the position of PNA binding. PNA **38** which is directed to the inverted repeat at bases 1540−1565 (Figure 3) was incubated with pUC19 at 37 °C for 5 min prior to addition of various conjugates, and the efficiency of priming by the conjugates was evaluated (Figure 10). Priming by conjugate **24**, which was complementary to the strand displaced by **38**, was unaffected. By contrast, the priming efficiency of conjugates **6** and **28** decreased dramatically in the presence of PNA **38**.

We performed similar experiments using PNA **36**, which is analogous in sequence to conjugate **22** (results not shown). As we had observed with PNA **38**, hybridization by PNA **36** reduced the priming efficiency of conjugates **6** and **28**. Addition of PNA **36**, however, did not reduce priming by conjugate **24**, even though the sites for hybridization of conjugates **6** and **28** were much farther away from the target for **36** than the site of hybridization for **24** (Figure 3, Tables 1 and 2). It is likely that hybridization of PNAs **36** and **38** were able to exert an effect at a distance by lowering the superhelical density of the plasmid, thereby reducing the propensity for strand invasion. Presumably, hybridization to the 1540−1565 inverted repeat is so favorable that it can be initiated even when prior hybridization at another site has removed some superhelical density, accounting for the maintenance of hybridization by **24** upon addition of **36**.

These results raise the possibility that spontaneous strand invasion can be used to alter recognition of distant DNA sequences as well as either promote or prevent recognition at their target sites. In our experiments action at a distance was facilitated by the circular nature of the plasmid, which prevents dissipation of topological information through rotation, but similar constraints also exist within the long linear chromosomes of eukaryotes, and domains of supercoiling within chromosomal DNA have been described.[34] Our results are in contrast to observations that the perturbation of strand displacement by homopyrimidine PNAs only propagates a few basepairs.[35] These experiments, however, used linear DNA, not supercoiled circular DNA, and would not have been expected to show distant effects.

Relevance of Hybridization Studies in a Cell-Free System to Targeting Strand Invasion Within Cells. In all of these experiments, annealing was performed under conditions of low ionic strength to encourage the formation of single-stranded structures and to promote hybridization. We note, however, that we have previously observed that hybridization is not affected by the presence of 70 mM sodium chloride or 70 mM potassium chloride and that hybridization can tolerate the presence of 1−2 mM magnesium chloride when excess DNA is added to lower the concentration of free cation.[15] In addition, hybridization of biotin-labeled PNAs survived exposure to 200 mM NaCl for 18 h during binding to streptavidin-coated beads (see Experimental Section), and PNAs have been reported to hybridize at triplet repeat-containing sequences within permeabilized cells[36] under conditions that approximate those of the physiological.

The situation within cells is likely to be further complicated by extremely high concentrations of protein and by factors that specifically promote dynamic fluctuations in DNA structure. Perhaps the most convincing evidence that strand invasion will be possible in vivo is provided by observations suggesting transiently single-stranded regions occur within cells.[26−29] Computational approaches have been developed to predict the location of such regions,[27] and experiment-based guidelines for the extrusion of small hairpins such as those in our studies have

(34) Kramer, P. R.; Sinden, R. R. *Biochemistry* **1997**, *36*, 3151−3158.
(35) Kurakin, A.; Larsen, H. J.; Nielsen, P. E. *Chem. Biol.* **1998**, *5*, 81−89.
(36) Boffa, L. C.; Morris, P. L.; Carpaneto, E. M.; Louissanit, M.; Allfrey, V. G. *J. Biol. Chem.* **1996**, *271*, 13228−13233.

recently been described.[28] In support of the physiological relevance of the potential for single-stranded regions is the fact that the regions within pUC19 that were most prone to strand invasion in our studies, bases 1540−1640 and 2540−2560, precede the origin of replication and the β-lactamase gene respectively (Figure 3).[30]

Conclusions

Our data suggest the following rules for strand invasion of duplex DNA by oligonucleotides: (i) Strand invasion is most efficient at sequences that both contain inverted repeats and are within AT-rich regions, but can also occur if only one of these characteristics is present. (ii) Hybridization by analogous PNAs and oligonucleotide−peptide conjugates is qualitatively similar. (iii) Hybridization exhibits a stringent dependence on sequence, with adjacent sequences varying dramatically in their propensity to bind oligonucleotides. (iv) PNA hybridization can either disrupt or promote hybridization of oligonucleotide−peptide conjugates, and these effects can be exerted over substantial differences. (v) Modification of PNAs with cationic amino acids can enhance hybridization, a result observed earlier for strand invasion at polypurine−polypyrimidine sequences.

Remarkably, strand invasion occurred at 13 of the 35 separate sequences chosen as hybridization targets. These 13 sequences cover over 11% of the plasmid, demonstrating the potential for strand invasion to be a general technique for recognition of duplex DNA. Furthermore, there is no reason that efficient hybridization should be confined to AT-rich regions or inverted repeats. Hybridization should also occur at other sequence motifs that tend to adopt non-B-type secondary structure or to the RNA polymerase open complex.[37]

The rules we describe will guide the targeting of oligonucleotides designed to control gene expression through binding to genomic DNA. Regions likely to assume structures that are partially single-stranded are common within promoters and other regions of DNA, making them targets for hybridization by chemically modified oligomers. Once bound, oligomers could act directly by blocking binding of transcription factors in a gene-specific fashion. Alternatively, our observation that DNA structure and function can be altered by hybridization at distant target sites suggests that they might also act through more subtle mechanisms to distort DNA structure and possibly upregulate or downregulate gene expression.

Control of gene expression by synthetic molecules is an important goal for chemical and biological science, and our approach emphasizes the potential of strand invasion by chemically modified oligomers for this purpose. Completion of the human genome project will provide basic information on approximately 100 000 genes. This immense store of data challenges chemists to further modify PNAs and other oligomers to improve the efficiency and generality strand invasion. Such versatile reagents would prove to be widely useful tools for dissecting gene function in complex systems and to be essential components of projects that aim to move from the one-dimensional understanding of protein function provided by sequence data to the multidimensional understanding needed to understand signaling pathways within cells.

Experimental Section

Synthesis of Oligonucleotide−Peptide Conjugates, PNAs, and PNA Derivatives. The synthesis of oligonucleotide−peptide conjugates was performed as described[15] by using disulfide exchange to conjugate 5′-*S*-thiopyridyl-containing oligonucleotides with peptides containing cysteine. The reagent for introducing a thiol at the 5′-termini of oligonucleotides was obtained from Clonetech (Palo Alto, CA). PNAs were obtained through automated synthesis using an Expedite 8909 synthesizer (Perkin-Elmer Biosystems, Foster City, CA) using the manufacturers protocols and were analyzed by mass spectral analysis as described.[38] PNA monomers and reagents were obtained from Perkin-Elmer Biosystems. Conjugates between PNAs and biotin or peptides were synthesized and analyzed as described.[39]

Strand Elongation by Oligonucleotide−Peptide Conjugates. Supercoiled plasmid pUC19 DN^{A30} ($\sigma = 0.5$) was prepared by a mild lysis protocol[40] followed by two successive CsCl gradient ultracentrifugations to minimize the likelihood of contamination by denatured or nicked duplex DNA. Inverted repeats were identified computationally using the program stemloop (Genetics Computer Group, WI). Sequences that were sensitive to nicking by S1 nuclease were identified using established protocols.[40] Hybridization of conjugate and plasmid was accomplished by mixing pUC19 (40 nM) with 20 equiv of oligonucleotide−peptide conjugate in 10 mM Tris-Cl, pH 7.5 buffer for 15 min at 37 °C. The hybridized primer−template mixture was then cooled on ice, and MgCl2, NaCl, and Tris-Cl, pH 7.5, were added to final concentrations of 8, 80, and 10 mM, respectively. The labeling mix consisting of modified T7 DNA polymerase (Sequenase, United States Biochemical, Cleveland, OH) (1 unit per reaction) and ^{35S}-dATP (Amersham) were added, and DNA sequencing using bound peptide−oligonucleotide as a primer was carried out. Equal volumes of the elongation reactions were applied to a denaturing 6% polyacrylamide gel and were separated by electrophoresis. The products were visualized by autoradiography and quantified using a Molecular Dynamics (Sunnyvale, CA) model 425F phosphorimager. To examine the effects of PNA hybridization on strand elongation by oligonucleotide−peptide conjugates, 5 equiv of PNA or PNA derivatives were mixed with pUC19 at a temperatures between 37 and 75 °C for varying periods of time prior to the addition of oligonucleotide−peptide conjugates. As a control, all conjugates were hybridized to denatured pUC19, and we observed that strand elongation was comparable for each conjugate regardless of sequence.

Affinity Capture of Biotin-Labeled PNA Conjugates. Affinity capture of plasmid DNA employed Dynabeads M-280 derivatized with streptavidin (Dynal, Oslo, Norway) as a matrix for separation of plasmids bound to biotin-labeled PNAs. pUC19 (40 nM) was mixed with biotin-labeled PNAs (200 nM) in 10 mM Tris-Cl pH 8.0 at 65 °C for 5 min prior to the addition of streptavidin-coated beads. Beads, plasmid, and PNA were incubated for 18 h at 22 °C in 10 mM Tris-Cl pH 8.0, 1 mM EDTA 200 mM NaCl. The beads were then washed with 10 mM Tris-Cl, 1 mM EDTA, 200 mM NaCl to remove unbound plasmid. Bound plasmid was eluted from the beads by incubating at 80 °C for 30 min in 10 mM Tris-Cl, 1 mM EDTA, 1 M NaCl. DNA in the supernatant was precipitated using ethanol and analyzed by 1% agarose gel electrophoresis.

Acknowledgment. The authors acknowledge Lynn Mayfield for expert assistance with PNA synthesis and purification. This work was supported by grants from the Robert A. Welch Foundation (I-1244), the National Institutes of Health (RO1-CA74908), and the Texas Advanced Technology Program (18603). T.I. is a Fellow with the Japanese Society for the Promotion of Science. D.R.C. is an Assistant Investigator with the Howard Hughes Medical Institute.

JA983834E

(37) The RNA polymerase open complex has already been targeted in cell-free systems using RNA and PNA oligomers. (a) Perrin, D. M.; Mazumber, A.; Sadeghi, F.; Sigman, D. S. *Biochemistry* **1994**, *33*, 3848−3854. (b) Larsen, H. J.; Nielsen, P. E. *Nucleic Acids Res.* **1996**, *24*, 458−463.

(38) (a) Mayfield, L. D.; Corey, D. R. *Anal. Biochem.* **1999**, in press. (b) Norton, J. C.; Waggenspack, J. H.; Varnum, E.; Corey, D. R. *Bioorg. Med. Chem.* **1995**, *3*, 437−445.

(39) Murchie, A. I. H.; Lilley, D. M. *Methods Enzymol.* **1992**, *211*, 158−180.

(40) (a) Panayotatos, N.; Wells, R. D. *Nature* **1981**, *289*, 466−470. (b) Lilley, D. M.; *Proc. Natl. Acad. Sci. U.S.A.* **1980**, *77*, 6468−6472.