Biochemistry (Moscow), Vol. 69, No. 8, 2004, pp. 855-860. Translated from Biokhimiya, Vol. 69, No. 8, 2004, pp. 1051-1058.

MINI-REVIEW

Regulation of Gene Expression via Triple Helical Formations

V. Karympalis, K. Kalopita, A. Zarros*, and H. Carageorgiou

Department of Experimental Pharmacology, Medical School, University of Athens, 75 Mikras Asias str., 11527 Athens, Greece; fax: 210-7462554; E-mail: azarros@med.uoa.gr

Received December 9, 2003
Revision received February 10, 2004

Abstract—Triplex-forming oligonucleotides (TFOs) are a challenging and very promising subject in modern biochemistry and molecular genetics. Over the last decade, several studies have stated that TFOs: a) can bind to DNA in a sequence-specific manner; b) can provoke DNA repair and recombination in mammalian cells, and c) can be a very effective biological tool in embryonic and oncogenic research on gene expression pathways. Herein we review the basic modes in which TFOs exist, their gene-regulation properties, as well as the obstacles that should be overcome before they become useful in clinical practice (chemotherapy and/or gene therapy).

Key words: triplex DNA, gene regulation, antisense and antigene drugs, chemotherapy

Techniques designed to regulate gene expression are of significant importance to the research fields of embryonic development, differentiation, experimental physiology, and pharmacology. Modifications in the expression of a single gene can alter levels of other genes and provide information on developmentally or metabolically regulated gene expression pathways. Moreover, handling the know-how of specific gene regulation could be very useful if applied for chemotherapeutic purposes.

Triplex DNA is the collinear association of three oligodeoxynucleotide (ODN) strands, usually occurring when a third strand binds to the major groove of a DNA double helix. Intramolecular triplexes form when a part of the DNA double helix folds back upon itself, whereas intermolecular triplexes involve the binding of a separate third strand [1]. Felsenfeld et al. [2] reported the first instance of triple helical nucleic acid structures in a study on the RNA system composed of poly(A) and poly(U). Forty-seven years later, triplex DNA is now known to have numerous potential applications as a molecular biological tool (such as manipulating genes by site-specific cleavage or mutagenesis in gene mapping and functioning studies). However, the potential of using triplex DNA as a gene modulator is of greater interest [3].

MODES OF TRIPLEX FORMATION

There have been many types of triplex DNA structures isolated so far, the most common of which are: a) intermolecular triplexes; b) intramolecular triplexes; c) peptide nucleic acid triplexes; d) recombination triplexes [4].

Intermolecular triplexes. Studies have shown that by using oligodeoxyribonucleotides, several groups developed triplex formation by pyrimidine-rich TFOs, which bind in the major groove of the cognate duplex in a parallel fashion [4]. The base pairing in the TA–T and C^+G–C triads showed no disruption of the Watson–Crick hydrogen bonds but afforded two additional Hoogsteen hydrogen bonds between the TFO and the purine-rich strand of the underlying duplex (see Figs. 1 and 2). Optimal target sequences must harbor consecutive purines on the same strand, since only purine bases are able to establish two Hoogsteen (or reverse Hoogsteen) type hydrogen bonds in the major groove of the DNA (which is the main restriction to the repertoire of potential target sites). Furthermore, cytosines must be protonated in order to form two hydrogen bonds with G (which is another important limitation). Recent research tries to [5]: a) increase triplex stability while preserving specificity; b) overcome the pH dependence in the (C, T) motif; c) extend the repertoire of potential target sequences.

Intramolecular triplexes. Oligopurine sequences presenting mirror symmetry can potentially adopt an intramolecular formation called H-DNA, involving both a triplex and a single-stranded structure. A likely *in vivo* role of H-DNA formation may involve regulation of transcription, since RNA polymerase creates regions of both negative and positive supercoiling. Antibodies specific for H-DNA can inhibit global transcription and replication processes on isolated nuclei or in permeabilized cells [6].

* To whom correspondence should be addressed.

0006-2979/04/6908-0855 ©2004 MAIK "Nauka/Interperiodica"

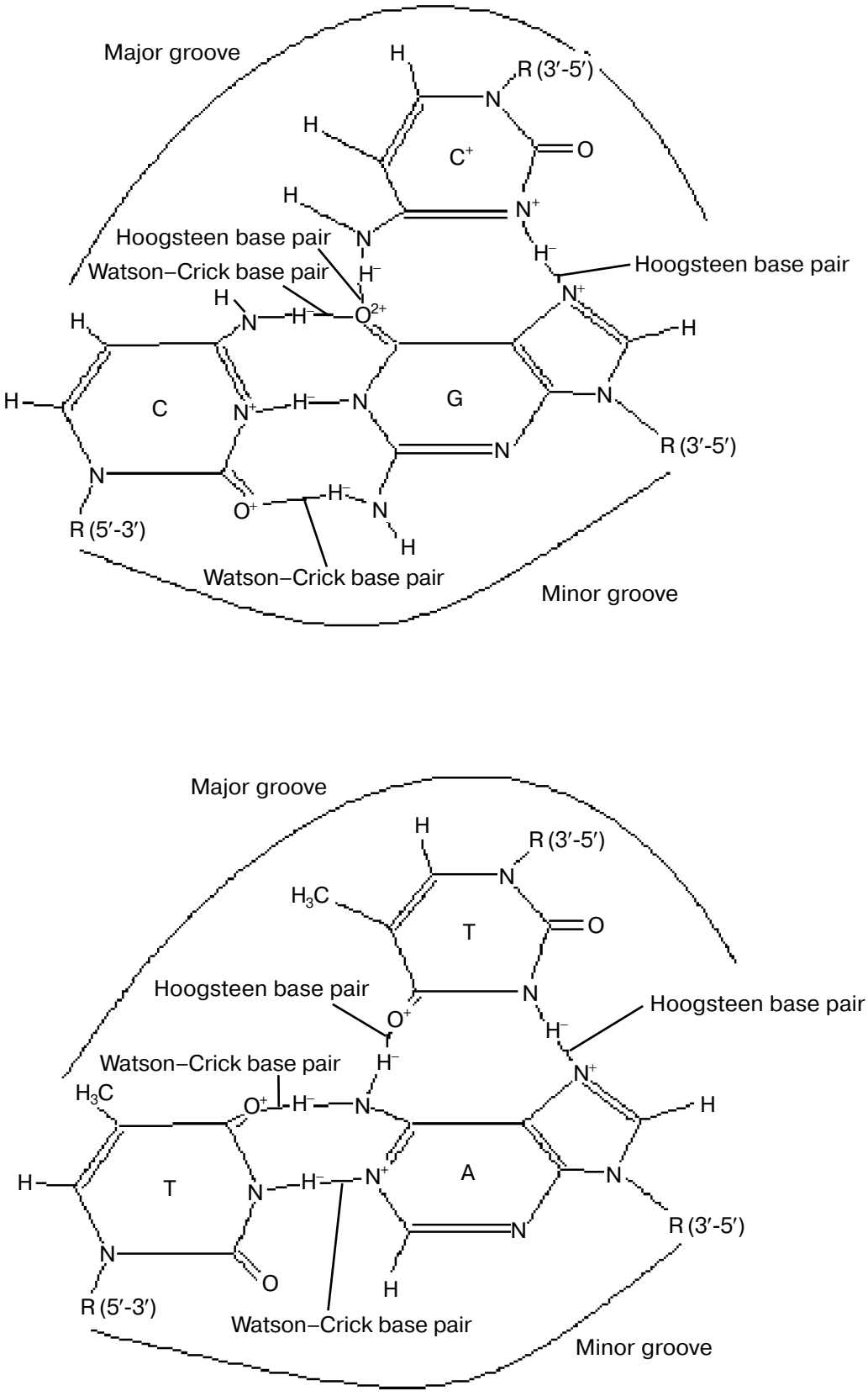

Fig. 1. Hoogsteen and Watson–Crick hydrogen bonding in triplex DNA structures.

However, direct proof is still lacking in ways to establish a correlation between H-DNA structure and gene expression within intact cells.

Peptide nucleic acid triplexes. Peptide nucleic acids (PNAs) are oligonucleotide analogs containing an uncharged polyamide backbone with otherwise normal DNA bases [7]. PNA oligomers form stable triplexes with only one of the DNA strands, leaving the complementary one displaced. This makes PNA very promising as an agent for sequence-specific binding with duplex DNA [8] and as a potential antigene drug [7, 9], since it is remarkably stable in biological fluids in which normal peptides and oligonucleotides are quickly degraded [10]. However, there are still some serious limitations considering its applications, that need to be by-passed (e.g., there is a dependency on ionic conditions for its possible sequence-specific targeting of DNA).

Recombination triplexes. Recombination DNA (R-DNA) triplexes require the RecA protein and therefore these triplexes are considered enzymic triplexes [11]. In R-DNA, the third strand is identical to one strand of the duplex and so binds parallel to that strand. A major difference with the above modes is that there is no restriction on the sequences capable of forming R-DNA triplexes (i.e., homopurine and homopyrimidine stretches). Studies recently gained special interest in the light of results reporting mutation correction by gene targeting in cell cultures using short double-stranded chimeric deoxy- and 2′-O-methyl oligonucleotides as the source of wild-type gene sequence. Although the mechanism of mutations is not yet definitely ascertained, repair is likely involved instead of recombination [12]. These results are encouraging and offer an alternative to the use of viral-based vectors in gene therapy, even though this strategy has met with some difficulties to be reproduced in different cell lines.

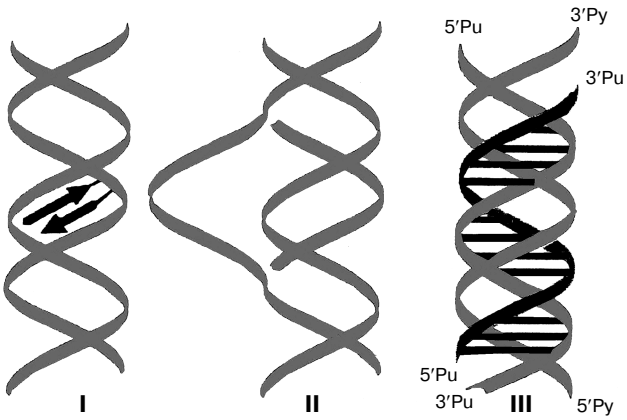

Fig. 2. Types of double-stranded DNA targeting: I) polyamide binding in the minor groove; II) PNA-induced strand displacement; III) TFO-binding through Hoogsteen or reverse-Hoogsteen base pairs.

MAKING USE OF TRIPLEX DNA STRUCTURES TO REGULATE GENE EXPRESSION

Due to the highly sequence-specific recognition of double-helical DNAs by TFOs, these molecules constitute a valuable tool for site-directed modification of genomic DNA. Synthetic TFOs have been designed with the purpose of using them to manipulate genes and gene function, both *in vitro* and *in vivo*. Their applications extend from inhibiting gene expression, by restricting protein binding to the DNA, and thus preventing its replication, to directing site-specific DNA damage and inducing mutagenesis, or even inducing DNA recombination, in conjunction with RecA proteins [4].

Transcription and replication inhibition: the antigene strategy. Triplex DNA structures can be designed to interfere with transcription and replication at different levels, as well as with the elongation step of the growing DNA and RNA molecules, by hampering DNA and RNA polymerase elongation procedures, respectively [13].

The first example of TFO-directed inhibition of gene transcription in cells was demonstrated by Postel et al. on the human *c-myc* gene, nearly twelve years ago. By targeting a site in the promoter region, the TFO presumably bound its target site on the duplex DNA, forming a triple helical DNA structure able to interfere with either the binding of transcription factors or the progression of the polymerase during transcription [14]. The mode of action is as follows: a) if the target site includes the recognition sequence for a transcription factor, then the TFOs compete with this factor for access to the major groove of the DNA; b) if the target site is close to the TATA box or between the TATA box and the transcription start, TFOs disrupt the assembly of the initiation complex. Finally, if the target site is in the transcribed region of the gene, triplex formation can cause premature termination of transcription [15]. In spite of all the above facts, there also seems to be a growing number of indications that TFOs can act as repressors of transcription in cell cultures as well, with the most convincing results so far being obtained for the interleukin-2 receptor *a* (IL-2Ra) promoter [16, 17]. Similar effects were also observed for human immunodeficiency virus (HIV) transcriptional inhibition in chronically infected cells [18].

The use of TFOs for DNA replication inhibition is less studied than their use for transcription inhibition. Only recently did biochemical scientists examine the effect of triplex formation on DNA replication. They demonstrated that TFOs could inhibit *in vivo* replication of DNA by either blocking the binding proteins, such as DNA polymerases (that are required for replication to the DNA), or by directly blocking the progression of the DNA polymerase along the DNA strand [19, 20].

Supporting the above, it was observed that *in vitro* formation of putative intramolecular triplexes (H-like triplexes) on single-stranded DNA templates trapped many different types of DNA polymerases, thus obstructing them from pursuing DNA replication [21]. However, despite the above claims, the conclusion that *in vivo* replication of DNA will be inhibited by triplex formation may be premature, because DNA helicase is undeterred in separating DNA strands in the presence of triplex structure, as Maine and Kodadek stated [22]. In addition to that, treatment of tissue culture cells with TFOs does not lead to cell cycle arrest, as it would be expected if DNA replication was halted.

It is understood that both transcription and replication inhibition induced by TFOs constitute the basis of an antigene strategy [23]. The basic idea is that the binding of a TFO to a target gene could prevent its normal functioning. Most studies of this strategy concerned the inhibition of transcription and were partially inspired by the existence of functionally important homopurine-homopyrimidine stretches in many eukaryotic promoters, which are appropriate targets for TFOs. Although there seems to be some discouragement about the *in vivo* inhibitory effects of TFOs, the antigene strategy, if used correctly, could potentially lead to rational drug design.

TFO-directed mutagenesis. Perhaps one of the most exciting prospects for triplex forming oligonucleotides is the possibility of generating site-specific mutations in mammalian cells, which provides a means to modify the genome in somatic cells by directly inactivating or correcting a gene of interest [24]. One such approach is to utilize triplex formation to induce site-specific DNA damage, by delivering DNA damaging agents to specific DNA sequences. Alkylating compounds, intercalators, photoactivatable cross-linkers, and ^{125}I have been attached to oligonucleotides used to form triplexes [25]. Furthermore, psoralen-modified TFOs have been employed to induce TA to AT transversions at a targeted site on extrachromosomal plasmids, both *in vitro* and *in vivo*, and on chromosomes in mammalian cells, such as the human adenosyl phosphoribosyl transferase gene (*Aprt*) [26], or the *supF* gene integrated in the genome of transgenic mice [27]. The damage seen includes single-strand nicking, double-stranded breaks, and photochemical cross-linking of the two duplex strands.

Under some circumstances, triplex formation alone appears to stimulate mutations in target genes, even when DNA damaging agents are not attached to the TFO. Noncovalent triplexes were shown to induce both point mutations and deletions more than 10-fold above the spontaneous mutation rate, by a process, which probably involves excision and transcription-dependant repair. These results were obtained on the *supF* gene with (G, A) TFOs [28] and more recently with short clamp-PNA [29], carried by a plasmid shuttle vector or a mouse chromosome, respectively. The triplex is apparently recognized as DNA damage and the damage repair pathways are activated. The "correction" of the perceived damage may itself cause the mutation through errors in repair. This point of view is also supported by the fact that TFO-induced mutagenesis was absent in cells deficient in the nucleotide excision repair (NER) damage recognition factor, XPA [28], suggesting that the triplex structure creates a sufficient helical distortion to provoke DNA repair, albeit in an error-prone manner. Similar results were observed in cells deficient in the CSB factor, which functions in the pathway of transcription-coupled repair [28].

The major problem with the use of TFOs for directed mutagenesis, as well as for transcription or replication inhibition, is in matching high sequence selectivity with binding that is sufficiently strong to interfere with genetic processes. Under physiological conditions, TFOs bind weakly to their targets, which by itself favors a high sequence selectivity. However, to significantly affect genetic processes, the TFO must be rather long, which limits the number of potential targets, as such long homopurine-homopyrimidine stretches are infrequent.

TFO-induced recombination. Homologous recombination is a general process in various organisms by which genetic information is rearranged to create genetic variations and which plays important roles in the repair of damaged DNA and is essential to reductive cell-division (meiosis) in eukaryotes. An apparatus for homologous recombination has to find homologous sequences within the entire genomic DNA to form intermolecular duplex points that are general intermediates for recombination. In 1979, it was discovered that the *Escherichia coli recA* gene product, RecA protein, recognizes and pairs homologous sequences between single- and double-stranded DNA to form stable heteroduplex joints, through a Mg^{2+}- and ATP-dependent reaction *in vitro* [30, 31]. This action, also known as the "synapsis step", requires searching for homology between the presynaptic filament and the double-stranded DNA target, which may be done by using Watson–Crick complementary rules on the partially separated double-stranded DNA forming a D-loop, or by triplex formation, creating what is known as "recombination", "parallel", or R-DNA [32]. The given names emphasize that chemically homologous DNA strands are parallel in R-DNA but antiparallel in standard triplexes and that any sequence can adopt an R-DNA conformation, while homopurine-homopyrimidine stretches are strongly preferable in adopting standard triplex structure.

Triplex induced recombination has been demonstrated in an SV40 shuttle vector in mammalian cells using psoralen-modified TFOs to damage the DNA, resulting in enhanced recombination frequencies of up to 25-fold [33], confirming the fact that double-stranded DNA breaks can stimulate recombination. Furthermore, similar to the results seen with TFO-directed mutagenesis, triplex formation alone is sufficient to stimulate

recombination in mammalian cells. When a plasmid containing duplicated copies of the *supF* reporter gene was transfected into mammalian cells in the absence of psoralen-conjunction, a fivefold induction in recombination occurred, which was found to be dependent on the presence of the NER pathway [34], engaged presumably to recognize the "lesion" and generate recombinogenic intermediates that contain some combination of either single- or double-stranded breaks.

In spite of the fact that the above data seem to be most consistent with the idea of "recombination" triplex formation, a careful analysis of three-stranded complexes formed under RecA protein using chemical probing indicates that base pairing in the parental duplex is disrupted. The incoming single-stranded DNA appears to form W–C pairs with the unwound complementary strand of the duplex, leading to the formation of D-loop-like structures, rather than "recombination" triplexes [35], a conclusion which is also supported by the fact that N7 position of guanines, involved in Hoogsteen hydrogen bonding, is not required for the formation of three-stranded complexes by RecA protein [36].

In conclusion, we must note that, even in the absence of a clear understanding of the structure of three-stranded joints promoted by RecA protein, interesting applications in gene targeting have already been found. The first example is the use of such complexes to block specific methylation sites in double-stranded DNA, with a technique called RARE, for RecA-Assisted Restriction Endonuclease cleavage, resulting in site-specific cleavage of DNA and separation of genome fragments, with obvious therapeutic applications [37].

APPLICATIONS OF TRIPLE HELICAL FORMATIONS

Although many advances have been made in triplex technology over the last decade, there are still limitations that need further investigation. Several parameters must be optimized to allow facile gene modification, including extending the triplex binding code, enhancing the stability of the oligonucleotide inside the cell, increasing the binding affinities of TFOs to their target duplexes, and many more. In several recent reports, site-specific psoralen-TFO adducts were detected in genomic DNA, using a single-strand ligation PCR assay [25]. But, while there is a wealth of indirect evidence for triplex formation on the chromosome (as indicated by triplex-directed transcription and replication inhibition, induced mutagenesis, and recombination), there is still very little direct physical evidence of triplex formation in intact cells. We herein describe the basic tools of antigene therapy, namely antisense and antigene drugs, and attempt a comparison of these methods with former pharmacological practices.

Antisense and antigene drugs. The principle of antisense drugs is simple: An oligonucleotide (mimic) is targeted to the mRNA of the gene exploiting nuclease complementarity and, upon binding of the antisense drug, the mRNA is inactivated and therefore does not direct the synthesis of the protein gene product. Inactivation can result from direct blockage of the translational machinery or from triggering of mRNA degradation by RNase H, an enzyme that specifically degrades the RNA strand of RNA/DNA duplexes. This makes necessary the development of a biologically stable oligonucleotide analog with superior hybridization capacity, commonly known as "antisense" oligonucleotide. Antigene drugs, on the other hand, focus on the fact that the genome can also act as a target for oligonucleotides and their analogs, because these may bind to homopurine or homopyrimidine sequences by means of triple-helix formation or, in the case of PNA, of strand invasion. Although *in vitro* experiments have indicated that gene transcription can be inhibited by oligonucleotides bound to promoter regions of the gene [17, 38] and by using PNAs, the synthesis of antigene drugs based on oligonucleotides and their analogs is probably even further into the future than the production of antisense drugs.

Pharmacology versus gene therapy. Gene therapy, involving the use of both antisense and antigene drugs, may prove an interesting approach in patient treatment in the future. DNA vectors can be constructed to generate RNA transcripts *in situ* in the nucleus, in such a way that the RNA sequence allows for triple helix formation on a targeted DNA sequence [39, 40]. This strategy was first tested to regulate insulin-like growth factor-1 (IGF-1) and its receptor gene at the transcriptional level [41], where only the RNAs containing the oligopurine sequence were shown to down-regulate IGF-1 and IGF-1R. Cloned cells down-regulated for the above genes were expended and re-injected into tumor-bearing animals, inducing tumor regression. Although these techniques appear to be very promising, there are still many questions to be answered and many obstacles to be surpassed [42] before they become part of everyday clinical practice and substitute modern pharmacological practice.

Many thanks are expressed to Dr. Achilleas Antonopoulos for his assistance in data accumulation.

REFERENCES

1. Scanlon, K. J., Ohta, Y., Ishida H., Kijima, H., Ohkawa, T., Kaminski, A., Tsai, J., Horng, G., and Kashani-Sabet, M. (1995) *FASEB J.*, **9**, 1288-1296.
2. Felsenfeld, G., Davies, D., and Rich, A. (1957) *J. Am. Chem. Soc.*, **79**, 2023-2024.
3. Praseuth, D., Guieysse, A. L., and Helene, C. (1999) *Biochim. Biophys. Acta*, **1489**, 181-206.

4. Vasquez, K. M., and Glazer, P. M. (2002) *Q. Rev. Biophys.*, **35**, 89-107.
5. Sun, J., Garestier, T., and Helene, C. (1996) *Curr. Opin. Struct. Biol.*, **6**, 327-333.
6. Agazie, Y. M., Burkholder, G. D., and Lee, J. S. (1996) *Biochem. J.*, **316**, 461-466.
7. Nielsen, P. E., Egholm, M., Berg, R. H., and Buchardt, O. (1993) *Anti-Cancer Drug Des.*, **8**, 53-63.
8. Demidov, V., Frank-Kamenetskii, M. D., Egholm, M., Buchardt, O., and Nielsen, P. E. (1993) *Nucleic Acids Res.*, **21**, 2103-2107.
9. 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., et al. (1992) *Science*, **258**, 1481-1485.
10. Demidov, V. V., Potaman, V. N., Frank-Kamenetskii, M. D., Egholm, M., Buchard, O., Sonnichsen, S. H., and Nielsen, P. E. (1994) *Biochem. Pharmacol.*, **48**, 1310-1313.
11. Camerini-Otero, R. D., and Hsieh, P. (1993) *Cell*, **73**, 217-223.
12. Ye, S., Cole-Strauss, A., Frank, B., and Kmiec, E. B. (1998) *Mol. Med. Today*, **4**, 431-437.
13. Hacia, J. C., Dervan, P. B., and Wold, B. J. (1994) *Biochemistry*, **33**, 6192-6200.
14. Postel, E. H., Flint, S. J., Kessler, D. J., and Hogan, M. E. (1991) *Proc. Natl. Acad. Sci. USA*, **88**, 8227-8231.
15. Dangle, J. M., and Weeks, D. L. (2001) *Differentiation*, **69**, 75-82.
16. Grigoriev, M., Praseuth, D., Robin, P., Hemar, A., Saison-Behmoaras, T., Dautry-Varsat, A., Thuong, N. T., Helene, C., and Harel-Bellan, A. (1992) *J. Biol. Chem.*, **267**, 3389-3395.
17. Grigoriev, M., Praseuth, D., Robin, P., Thuong, N. T., Helene, C., and Harel-Bellan, A. (1993) *Proc. Natl. Acad. Sci. USA*, **90**, 3501-3505.
18. McSchan, W. M., Rossen, R. D., Laughter, A. H., Trial, J., Kessler, D. J., Zendegui, J. G., Hogan, M. E., and Orson, F. M. (1992) *J. Biol. Chem.*, **267**, 5712-5721.
19. Birg, F., Praseuth, D., Zerial, A., Thuong, N. T., Asseline, U., Le Doan, T., and Helene, C. (1990) *Nucleic Acids Res.*, **18**, 2901-2908.
20. Volkmann, S., Jendis, J., Frauendorf, A., and Moelling, K. (1995) *Nucleic Acids Res.*, **23**, 1204-1212.
21. Samadashwily, G. M., Dayn, A., and Mirkin, S. M. (1993) *EMBO J.*, **12**, 4975-4983.
22. Maine, I. P., and Kodadek, T. (1994) *Biochem. Biophys. Res. Commun.*, **204**, 1119-1124.
23. Helene, C. (1991) *Anticancer Drug Res.*, **6**, 569-584.
24. Vasquez, K. M., Wang, G., Havre, P., and Glazer, P. M. (1999) *Nucleic Acids Res.*, **27**, 1176-1181.
25. Giovannangeli, C., and Helene, C. (1997) *Antisense Nucleic Acid Drug Dev.*, **7**, 413-421.
26. Vasquez, K. M., Wensel, T. G., Hogan, M. E., and Wilson, J. H. (1996) *Biochemistry*, **35**, 10712-10719.
27. Gunther, E., Havre, P., Gasparro, F., and Glazer, P. M. (1996) *Photochem. Photobiol.*, **63**, 207-212.
28. Wang, G., Seidman, M. M., and Glazer, P. M. (1996) *Science*, **271**, 802-805.
29. Faruqi, A., Egholm, M., and Glazer, P. (1998) *Proc. Natl. Acad. Sci. USA*, **95**, 1398-1403.
30. Shibata, T., DasGupta, C., Cunningham, R. P., and Radding, C. M. (1979) *Proc. Natl. Acad. Sci. USA*, **76**, 1638-1642.
31. McEntee, K., Weinstock, G. M., and Lehman, I. R. (1979) *Proc. Natl. Acad. Sci. USA*, **76**, 2615-2619.
32. Zhurkin, V. B., Raghunathan, G., Ulyanov, N. B., Camerini-Otero, R. D., and Jernigan, R. L. (1994) *J. Mol. Biol.*, **239**, 181-200.
33. Faruqi, A. F., Seidmann, M. M., Segal, D. J., Carroll, D., and Glazer, P. M. (1996) *Mol. Cell Biol.*, **16**, 6820-6828.
34. Faruqi, A. F., Datta, H. J., Carroll, D., Seidman, M. M., and Glazer, P. M. (2000) *Mol. Cell Biol.*, **20**, 990-1000.
35. Adzuma, K. (1992) *Genes Dev.*, **6**, 1679-1694.
36. Jain, S. K., Inman, R. B., and Cox, M. M. (1992) *J. Biol. Chem.*, **267**, 4215-4222.
37. Ferrin, L. J., and Camerini-Otero, R. D. (1991) *Science*, **254**, 1494-1497.
38. Duval-Valentin, G., Thuong, N., and Helene, C. (1992) *Proc. Natl. Acad. Sci. USA*, **89**, 504-508.
39. Ilves, H., Barske, C., Junker, U., Bohnlein, E., and Veres, G. (1996) *Gene*, **171**, 203-208.
40. Ohno, M., Fukagawa, T., Lee, J. S., and Ikemura, T. (2002) *Chromosoma*, **111**, 201-213.
41. Rininsland, F., Johnson, T. R., Chernicky, C. L., Schulze, E., Burfeind, P., and Ilan, J. (1997) *Proc. Natl. Acad. Sci. USA*, **94**, 5854-5859.
42. Seidman, M. M., and Galzer, P. M. (2003) *J. Clin. Invest.*, **112**, 487-494.