Pergamon

Tetrahedron, Vol. 53, No. 35, pp. 12041-12056, 1997

0040-4020/97 $17.00 + 0.00

PII: S0040-4020(97)00767-9

Selective Base-Pair Destabilization Enhances Binding of a DNA Methyltransferase

Daniel A. Erlanson, Scot A. Wolfe, Lin Chen and Gregory L. Verdine*

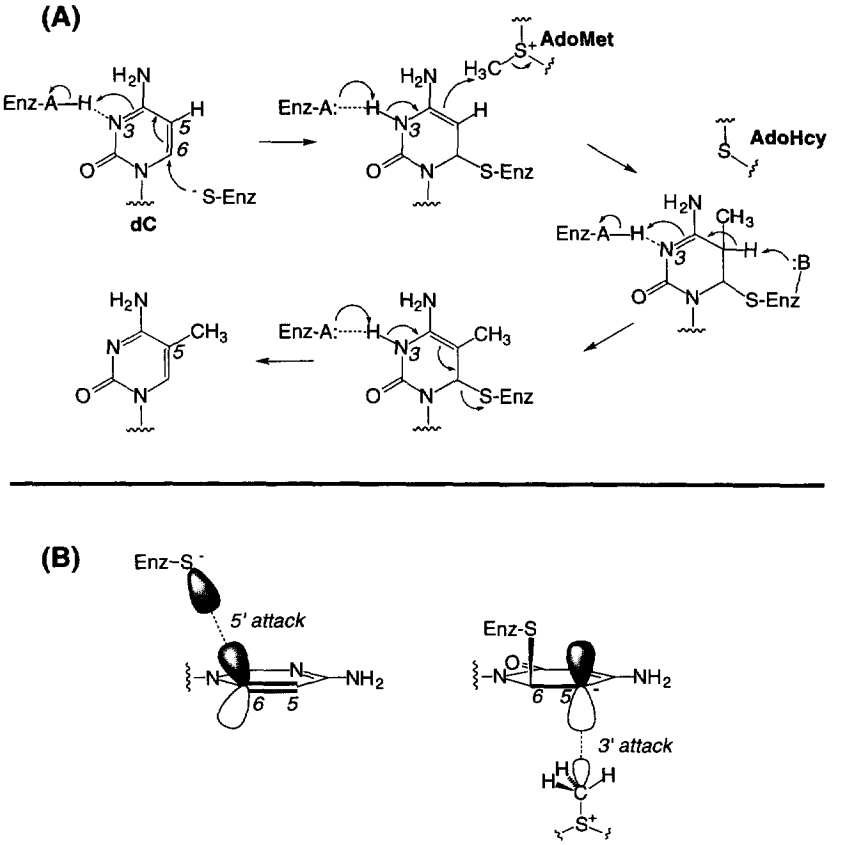
Harvard Department of Chemistry and Chemical Biology
12 Oxford St., Cambridge, MA 02138, USA

Abstract: Disulfide crosslinks were introduced into the minor groove of DNA using the convertible nucleoside approach. Depending upon the length of the tether, the modified base pairs were either stabilized or destabilized. When the base-pairs were destabilized, the oligonucleotide was bound by a DNA (cytosine-5)-methyltransferase (DCMtase) about 50-fold more tightly than the corresponding unmodified oligonucleotide. Insights into the mechanism of DCMtases are discussed.

INTRODUCTION

DNA methylation is an essential biological process that has been studied for decades,[1] yet continues to pose fascinating scientific challenges.[2,3] Bacteria express methyltransferases (MTases) to protect their own genomes against restriction endonucleases, which cleave the DNA of invading viral pathogens.[4] Eukaryotes also produce MTases, and the methylation of cytosines in the mammalian genome has been implicated in a variety of biological processes. For example, genomic imprinting, in which inherited alleles of the same gene are expressed at unequal dosages, has been traced to differential methylation of the paternal and maternal alleles.[5] Furthermore, normal development in mice is dependent on the presence of a functional MTase gene,[6] and the expression levels of several proteins have been correlated with the extent of methylation of their genes.[7] Indeed, aberrant methylation has been implicated in both tumorigenesis[8] and in Fragile X syndrome,[9] the leading cause of hereditary mental retardation. The recent discovery that unmethylated CpG motifs trigger B-cell activation in vertebrates underscores the long-standing evolutionary importance of eukaryotic DNA methylation.[10] Furthermore, since deamination of 5-methylcytidine to thymidine is more rapid than deamination of cytidine to uracil, methylation increases the mutagenic burden on the genome.[11]

The mechanism of enzyme-mediated cytosine methylation by DNA (cytosine-5)-methyltransferases (DCMtases) involves conjugate addition of a cysteine thiolate to the carbon 6 (C6) of the substrate cytosine followed by transfer of a methyl group to C5. β-elimination then regenerates the free enzyme (Scheme 1A).[12,13] Because the stereoelectronic attack trajectories for thiolate addition and methyl transfer cannot be accommodated in B-form DNA (Scheme 1B), we proposed that a transient helical disruption must occur during the catalytic event.[14] Disruption of base-pairing would also make N3 available as a proton acceptor, thus avoiding the energetically unfavorable step of forming a carbanion intermediate. Recent crystallographic structures of DCMtases bound to DNA have illuminated the precise nature of this transient helical disruption by showing that the substrate cytosine is fully everted from the DNA helix and is inserted into an active site pocket in the enzyme.[15,16] Remarkably, in one of these structures, eversion of the substrate cytosine is accompanied by the formation of frame-shifted base-pairs.[16] What remains uncertain, however, is the mechanism by which the substrate DNA undergoes such radical deformation.

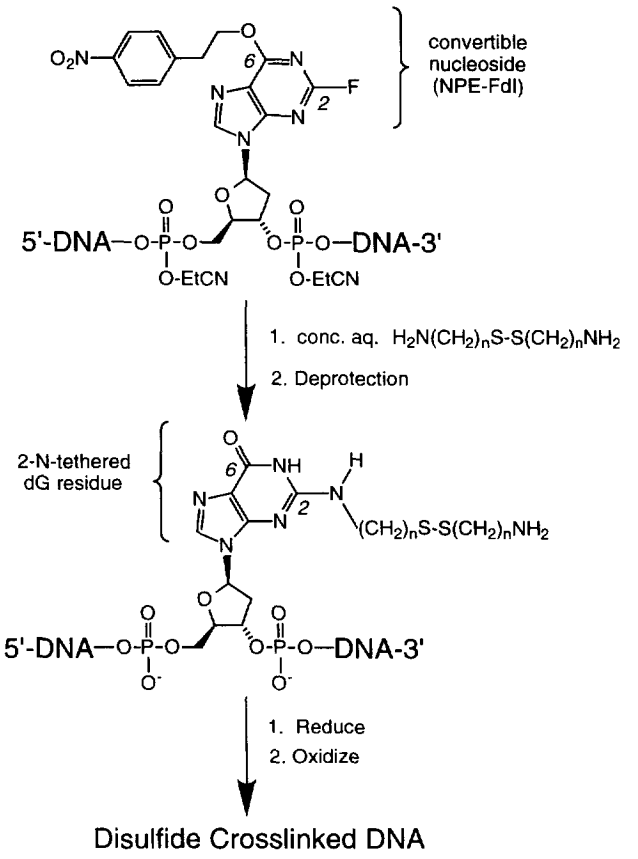


Scheme 1. (A) Proposed mechanism of DCMtase mediated cytosine methylation.[12,13] (B) Attack trajectories of the thiolate addition and cytosine methylation.[2]

The ability to engineer disulfide crosslinks into nucleic acids has proven valuable in the analysis of protein-DNA and protein-RNA interactions.[17] Although disulfide bonds were reported to occur in bacterial tRNA as early as 1967,[18] it was not until much later that chemistry was developed that permitted the site-specific introduction of disulfide crosslinks into synthetic DNA and RNA.[19] These crosslinks stabilized the duplex structure, and demonstrated that disulfide crosslinking of DNA could be a powerful technique for modifying the properties of nucleic acids.[14,17,20-36] Subsequently, we and others have found that disulfide tethers can stabilize DNA and RNA hairpins and other structures[20,26,29,31], that a strained disulfide tether can introduce sufficient torsional stress into duplex DNA to disrupt Watson-Crick base-pairing,[24] and that disulfide tethers can be used as probes for protein-induced DNA distortion.[14,25,29,29,30] Here we report the synthesis of disulfide-crosslinked DNA with either enhanced or decreased base-pair stability and its use as a probe of the method by which a DCMtase gains access to its target cytosine.

RESULTS AND DISCUSSION

In a previous communication,[14] we reported the synthesis of DNA modified in the minor groove using a post-synthetic procedure known as the convertible nucleoside approach.[37] The optimized procedure is briefly described here (Scheme 2). A self-complementary oligonucleotide containing the 5'-GGCC-3' recognition site for M. HaeIII DCMtase was synthesized using solid phase synthesis. Initial biochemical studies were performed using 18-mers; detailed structural characterization was done on decamers. The convertible nucleoside O^6-[2-(*p*-Nitrophenyl)ethyl]-2-fluoro-2'-deoxyinosine (NPE-FdI) was incorporated into the appropriate position during solid-phase synthesis. The resin-bound oligonucleotide was treated with concentrated aqueous solutions of disulfide diamines ($[H_2N(CH_2)_nS]_2$ n = 2,3) or 2-(methylthio)ethylamine (MTE amine), resulting in replacement of the 2-fluoro substituent by the respective N-alkyl disulfide or N-MTE group. The oligonucleotides were then treated with 1,8-diazabicyclo[5.4.0]undec-7-ene (DBU) in formamide to remove the NPE protecting group. The disulfide-containing self-complementary oligonucleotides were reduced by treatment with dithiothreitol, purified by reversed phase HPLC, and then oxidized to provide cross-linked decamers and octadecamers **C_2X-10/18** and **C_3X-10/18** in high yield (Figure 1). In the case of the 18-mers, air oxidation of the free thiols to the disulfide proved to be efficient; for the shorter 10-mers we used 5,5'-dithio-bis(2-nitrobenzoic acid) (DTNB)[38] as described previously.[30]



Scheme 2. Synthesis of disulfide crosslinked DNA.

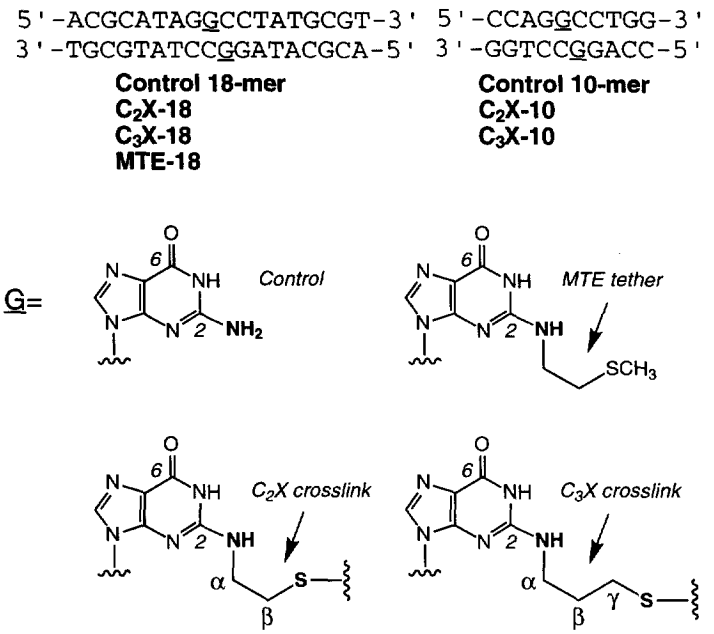


Figure 1. Oligonucleotides described in the text.

To verify that the oligonucleotides contained a disulfide crosslink, they were analyzed by nucleoside composition analysis, representative examples of which are shown in Figure 2. The nucleoside composition analysis of **C_2X-10** revealed a single new peak (Figure 2A) with a UV spectrum similar to deoxyguanosine (dG) and a mass spectrum (not shown) consistent with a disulfide cross-linked dimer of dG. Treatment of the digested oligonucleotide with the reducing agent tris(2-carboxyethyl)phosphine (TCEP) led to the rapid reduction of the new peak to an earlier-eluting species (Figure 2B) with a similar UV profile. Similar results were observed for the other disulfide containing oligonucleotides.

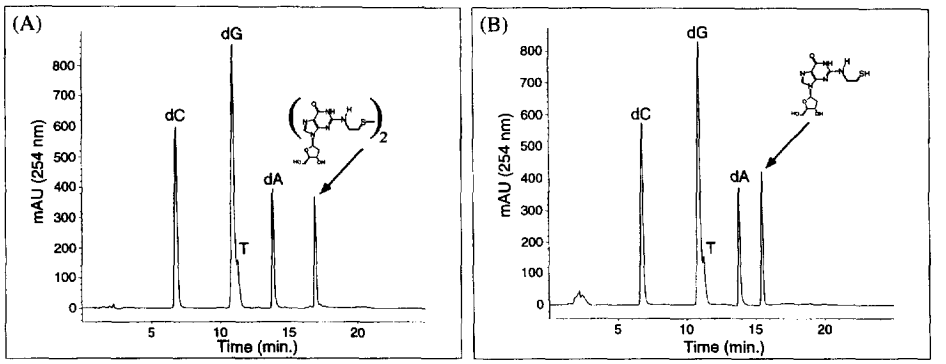


Figure 2. (A) Nucleoside composition analysis of **C_2X-10**. (B) Nucleoside composition analysis of **C_2X-10** after reduction with TCEP.

Before high resolution structural information was available on any DCMtase-DNA complex, the initial goal of these experiments was to determine whether disulfide-crosslinking would affect DCMtase activity. Because of the stereoelectronic arguments presented above, we reasoned that distortion of the duplex around the substrate cytosine would be necessary for the reaction to occur. Were the enzyme able to process a disulfide crosslinked substrate this would suggest that the enzyme does not melt a large stretch of the DNA, but rather effects a highly localized remodeling of duplex structure.

To assess the effect of a minor groove DNA cross-link on recognition by the DCMtase M. HaeIII, we carried out equilibrium binding assays under noncatalytic conditions. Binding of M. HaeIII to the cross-linked duplexes was compared with binding of unmodified DNA and **MTE-18** (Figure 1). Wild-type M. HaeIII bound **MTE-18** and $\mathbf{C_3X}$**-18** with affinity similar to that of the unmodified oligonucleotide (Table 1). In contrast, $\mathbf{C_2X}$**-18** was bound by M. HaeIII nearly 50-fold more tightly than unmodified DNA (Figure 3 and Table 1). These effects do not depend upon the ability of the enzyme to form a covalent complex with the DNA, since similar results were obtained with a mutant M. HaeIII in which the active site cysteine was replaced by alanine.[13,39]

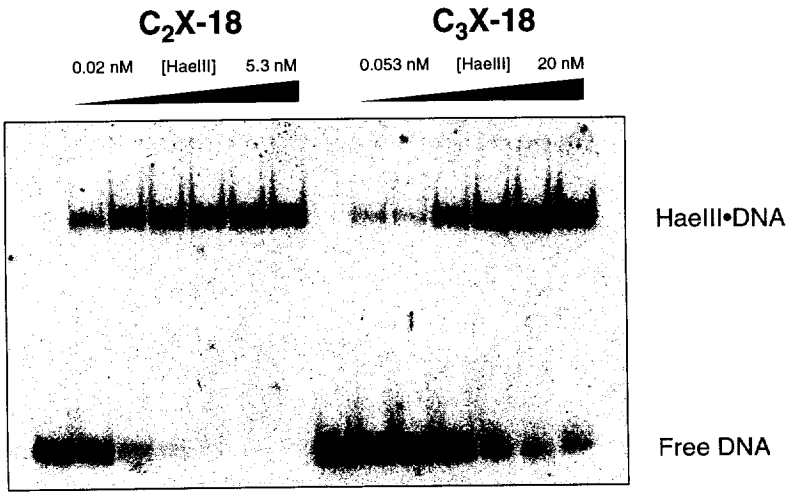


Figure 3. EMSA of $\mathbf{C_2X}$**-18** (**3a**, lanes 1 - 7) and $\mathbf{C_3X}$**-18** cross-link (**3b**, lanes 8 - 14). The labeled probe concentration in all reactions was 0.035 nM. The concentration of M HaeIII DCMtase was 0, 0.020, 0.053, 0.20, 0.53, 2.0, and 5.3 nM (lanes 1 - 7) and 0, 0.053, 0.20, 0.53, 2.0, 5.3, and 20.0 nM (lanes 8 - 14).

In methyltransferase experiments conducted under conditions of substrate saturation, both cross-linked molecules were found to be substrates for M. HaeIII, albeit less efficient ones than the control.[14] Indeed, the two crosslinked substrates were methylated at comparable rates to the uncrosslinked **MTE-18** control, suggesting that the reduced rate of DNA methylation relative to the unmethylated control can be ascribed to attachment of the tether rather than crosslinking *per se*.[40] The fact that crosslinked DNA serves as a substrate for

M. HaeIII suggested that global strand separation is not necessary for catalysis.[14] This prediction has been confirmed crystallographically.[15,16]

To address the origin of the unusually strong binding of M. HaeIII for **C_2X-18**, we measured the kinetics of dissociation of the enzyme•substrate complex. These studies were designed to determine the rate constant k_{off} in the equilibrium:

$$\text{M. HaeIII•DNA} \underset{k_{on}}{\overset{k_{off}}{\rightleftharpoons}} \text{M. HaeIII + DNA}$$

$$K_D = k_{off}/k_{on}$$

As seen in Table 1, M. HaeIII dissociates from **C_2X-18** about 10 times more slowly than from **C_3X-18**. This, however, does not account for the total reduction in K_D for the M. HaeIII•**C_2X-18** complex, suggesting that k_{on} may also be increased by a factor of 5. The rate constants k_{on} and k_{off} measure different properties of the enzyme-DNA complex. A lower k_{off} reflects the energy barrier that M. HaeIII must overcome in dissociating from the substrate, which could be increased if the extrahelical substrate cytosine is more stable in **C_2X-18** than in **C_3X-18**; an increased k_{on} reflects "pre-distortion" of the DNA which might allow the enzyme to bind to its substrate more rapidly. This distortion, in turn, could reflect either a single conformer preferred by the enzyme or a fluxional state in which the enzyme-preferred conformation (perhaps the transient unpairing of the substrate cytosine) may occur more frequently in **C_2X-18** than in **C_3X-18** or in unmodified DNA.

Oligonucleotide	K_D (relative)	k_{off} (min^{-1})
Control 18-mer	43 ± 29	n.d.
MTE-18	85 ± 44	n.d.
C_2X-18	1	$1.4 \pm 0.6 \times 10^{-3}$
C_3X-18	55 ± 20	$1.5 \pm 0.2 \times 10^{-2}$

Table 1. Equilibrium dissociation constants (K_D) and dissociation rate constants (k_{off}) for oligonucleotide•HaeIII complexes. K_D values are normalized to the K_D of **C_2X-18**; a relative K_D of 1 corresponds to approximately 0.013 nM.

To gain insight into the relative thermodynamic stabilities of **C_2X-18** and **C_3X-18**, we determined their susceptibility to reduction. We anticipated that if the disulfide bond is relatively unstrained, it should be thermodynamically stable and therefore relatively resistant to reduction. On the other hand, if the tether forces the oligonucleotide to adopt an energetically unfavorable conformation, we would expect this strain energy to manifest itself as an increase in the susceptibility of the disulfide bond to reduction.[30] As is seen in Figure 4, **C_2X-18** is dramatically more sensitive to reduction by 2-mercaptoethanol (β-ME) than is **C_3X-18**. Whereas **C_3X-18** is stable even at 100 mM β-ME for several hours, **C_2X-18** is almost completely reduced under the same conditions.

This difference in susceptibility to reduction could be attributed to two phenomena. First, it is possible that the actual reduction potential is different for the two oligonucleotides. It is well-established that

conformational effects can thermodynamically effect the redox potential of a disulfide in model systems as well as in proteins.[41-43] The other explanation is kinetic: perhaps the disulfide bond in **C_3X-18** is less accessible than that in **C_2X-18**, and is therefore shielded from the reductant. In order to distinguish between these two possibilities, we performed kinetic analyses using the irreversible reducing agent tris-2-carboxyethylphosphine (TCEP).[44,45] We determined the first-order rate constants for the reduction reactions to be 310 ± 50 $hr^{-1}M^{-1}$ for **C_2X-18** and 146 ± 47 $hr^{-1}M^{-1}$ for **C_3X-18**; that is, the oligonucleotide with the shorter tether is reduced about twice as rapidly as the oligonucleotide with the longer tether. This minor difference in the kinetic susceptibility of the two disulfide tethers suggests that their differential reactivity is due largely to the relative thermodynamic instability of the shorter crosslink.

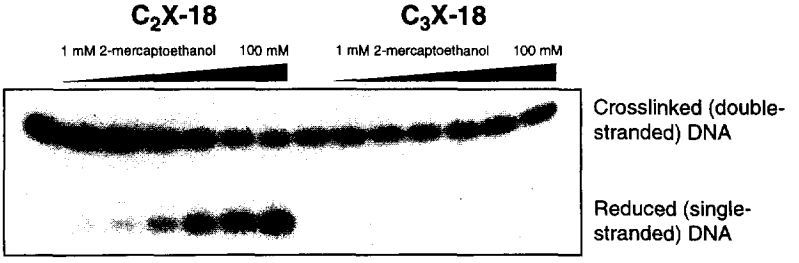


Figure 4. Susceptibility of **C_2X-18** and **C_3X-18** to reduction by 2-mercaptoethanol. The concentrations of β-ME are 0, 1, 3, 10, 25, 50, and 100 mM.

To obtain a more detailed understanding of the local structural and dynamic perturbations of the disulfide tethers, we synthesized two cross-linked decamers: **C_2X-10** and **C_3X-10**. We chose to work with these decameric oligonucleotides because the corresponding unmodified decamer has been extensively studied by both NMR spectroscopy[24] and crystallography.[46] We initially characterized these three oligonucleotides by examining their response to thermal denaturation.[47] Surprisingly, neither of the crosslinked molecules showed a typical cooperative melting transition, even at temperatures as high as 95 °C. In contrast, the unmodified control decamer melted at 55.7 ± 0.6 °C. We further investigated the apparent lack of a melting transition in the crosslinked molecules using circular dichroism (CD) spectroscopy, a technique that is frequently utilized to examine the global conformational properties of DNA.[48,49] The CD spectra of **C_2X-10**, **C_3X-10**, and the unmodified decamer are shown in Figure 5. At 7 °C, the spectra are qualitatively similar, indicating that the tethers do not induce dramatic conformational changes. Furthermore, the spectra suggest that all three molecules are B-form. However, at 80 °C, the unmodified decamer shows a weak CD signal, while both crosslinked oligonucleotides still retain substantial duplex structure. These data suggest that the duplex DNA is dramatically stabilized by the presence of either tether. Global stabilization attributable to the disulfide crosslink was also observed for **C_2X-18** and **C_3X-18**, which melt at 86.3 °C and 90.1 °C, respectively, compared to 71.3 °C and 73.2 °C for the unmodified 18-mer and **MTE-18**.[14] It is interesting that the 18-mers all show cooperative melting transitions while the decamers do not. We propose that the tether enforces local structure at the site of the crosslink even at high temperatures; because this core is proportionally larger for a decamer than an 18-mer, the melting curve does not show a clear inflection point. In the case of the 18-mer, even though a small portion

of the molecule may retain structure, the duplex as a whole denatures, and therefore a sharper transition is observed.

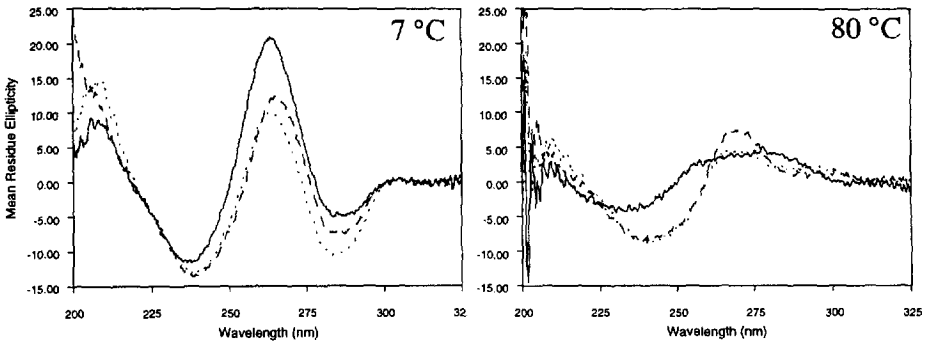


Figure 5. CD spectra of control decamer (solid), C_3X-10 (dashed), and C_2X-10 (dotted).

We set out to establish detailed structural information through nuclear magnetic resonance spectroscopy. Two-dimensional ^{1}H-^{1}H NOESY, COSY, and TOCSY spectra[50-52] were used to assign the majority of the protons and phosphorous atoms in the three oligonucleotides, which are presented in Table 2. The chemical shifts of the nonexchangeable protons in all three oligonucleotides are quite similar, except in the immediate vicinity of the tether. Most of the significant deviations from the control oligonucleotide occur at the 1' or 4' protons (in the minor groove) of bases 5-7. These could be caused by small structural perturbations, increased stiffness, or trivially by the physical proximity of the tether.

The C1' to base region from the ^{1}H-^{1}H NOESY for C_3X-10 is shown in Figure 6. Connectivity typical of B-form DNA is observed for all three oligonucleotides in both the H1' to base regions and H2'/2" to base regions of their NOESY spectra.[50] These results, along with the CD spectra, indicate that neither tether changes the global conformation of the DNA.

Close examination of the COSY and NOESY spectra allowed assignment of the tether protons. The two dimensional spectra of C_3X-10 reveal that all six tether protons are diastereotopic (Table 2). In contrast, the β methylene protons in the C_2X-10 tether are degenerate, suggesting that this tether is in rapid conformation equilibration on the NMR timescale (Table 2). If the tether were static in the minor groove, we would anticipate dissimilar chemical shifts for these protons, since their chemical environments would be different.

Regarding the physical environment of the crosslink, the tether in C_3X-10 shows extensive NOE contacts to the 1' protons of C6 and C7, as well as the to base-paired amino proton of G5. Fewer and weaker contacts are also observed to the 1' proton of G5, the 4' proton of C7, and the imino proton of G5. These NOE crosspeaks are expected based on molecular modeling exercises. The tether in C_2X-10 shows some NOE contacts to the 1' protons of cytidine 6 (C6) and cytidine 7 (C7) as well as to the amino proton of guanosine (G5). The weaker intensities of these NOE crosspeaks are consistent with an increased distance between the tether and the backbone, expected because of the shorter length of this tether. These weaker NOE crosspeaks could also be caused by conformational averaging on the NMR timescale.

Control 10-mer

Base	H8/H6	H5/H2/Methyl	H1'	H2'	H2''	H3'	H4'	Imino	Amino	Phosphate
C1	7.51	5.68	5.70	1.79	2.23	4.41	3.87	-	6.80/7.71	-4.08
C2	7.36	5.46	4.95	1.91	2.11	4.59	3.84	-	6.82/8.44	-3.79
A3	7.97	7.55	5.77	2.55	2.68	4.83	4.17	-	6.21/7.80	-4.19
G4	7.47	-	5.33	2.35	2.43	4.77	4.15	12.68		-4.02
G5	7.44	-	5.65	2.30	2.46	4.69	4.17	12.73		-4.23
C6	7.13	4.97	5.71	1.82	2.27	4.55	3.88	-	6.18/7.82	-4.23
C7	7.31	5.29	5.67	1.94	2.20	4.58	3.99	-	6.72/8.18	-4.38
T8	7.16	1.43	5.33	1.79	2.03	4.59	3.89	13.97	-	-4.02
G9	7.63	-	5.44	2.45	2.50	4.75	4.14	12.90		-3.93
G10	7.61	-	5.90	2.32	2.10	4.43	3.99	13.02		

C_3X-10

Base	H8/H6	H5/H2/Methyl	H1'	H2'	H2''	H3'	H4'	Imino	Amino	Phosphate
C1	7.47	5.62	5.65	1.73	2.20	4.39	3.84	-	6.75/7.67	-4.10
C2	7.35	5.43	4.88	1.87	2.09	4.57	3.81	-	6.80/8.39	-3.78
A3	7.91	7.62	5.75	2.51	2.64	4.82	4.15	-	6.24/7.82	-4.16
G4	7.54	-	5.20	2.40	2.40	4.77	4.14	12.64	6.49/8.09	-3.91
G5	7.44	-	5.83	2.29	2.50	4.76	4.20	12.75	-/8.77	-4.37
C6	6.97	**4.77**	**5.97**	1.69	2.28	4.62	3.84	-	**5.82/7.55**	-4.19
C7	7.26	5.23	5.63	**1.71**	2.14	4.62	**3.78**	-	6.64/8.22	-4.26
T8	7.18	1.43	**5.11**	1.86	2.00	4.56	3.83	13.99	-	**-3.80**
G9	7.61	-	5.45	2.48	2.50	4.75	4.13	12.93		-3.98
G10	7.60	-	5.87	2.34	2.09	4.43	3.98	13.00		

N2-propanethiol tether	α	β	γ
	2.77/3.08	1.62/1.87	2.50/2.66

C_2X-10

Base	H8/H6	H5/H2/Methyl	H1'	H2'	H2''	H3'	H4'	Imino	Amino	Phosphate
C1	7.43	5.55	5.60	1.72	2.18	4.35	3.83	-	6.67/7.73	-4.07
C2	7.34	5.42	4.87	1.87	2.06	4.57	3.82	-	6.78/8.41	-3.75
A3	7.96	7.57	5.69	2.54	2.64	4.81	4.16	-	6.26/7.78	-4.16
G4	7.46	-	5.38	2.32	2.46	4.76	4.16	12.68		-4.21
G5	7.39	-	5.79	**1.98**	2.36	4.70	4.03	**11.79**	-/8.03	-4.21
C6	7.10	5.13	5.87	1.85	2.17	4.58	3.79	-	**6.41/7.08**	-4.07
C7	7.31	5.20	5.69	1.82	2.16	4.57	3.88	-	**6.46**/8.10	-4.24
T8	7.14	1.44	5.28	1.81	2.01	4.57	3.85	13.93	-	-3.92
G9	7.62	-	5.41	2.46	2.49	4.75	4.13	12.90		-3.95
G10	7.57	-	5.85	2.32	2.10	4.43	3.98			

N2-ethanethiol tether	α	β
	2.52/2.64	3.14

Table 2. Chemical shifts of protons and phosphorous atoms in the three decamers. All chemical shifts are in parts per million (ppm) and are referenced to HDO (4.67 ppm) or trimethyl phosphate (0.0 ppm). Chemical shifts which differ by more than 0.2 ppm between the control decamer and the crosslinked decamer are shown in bold. All spectra were recorded at 7 °C.

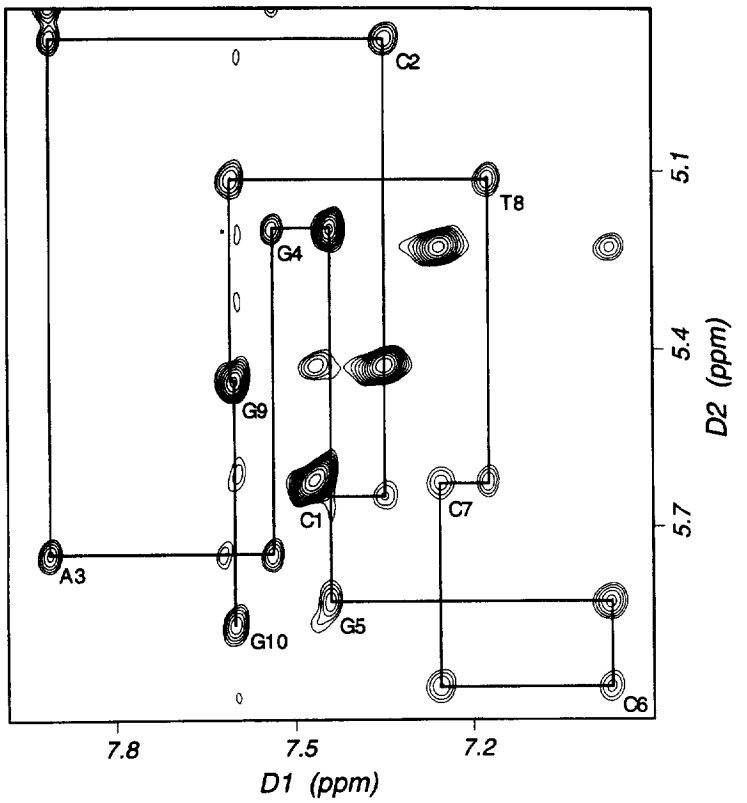


Figure 6. Two dimensional ^{1}H-^{1}H NOESY spectrum of **C_3X-10** showing the base to 1' region. The sequential connectivity of NOE crosspeaks observed between the base H6/H8 and its H1' proton indicates that the oligonucleotide is B-form. Note that the C1 H6 to 1' crosspeak is obscured beneath the H6-H5 crosspeak.

One-dimensional NMR in 10% D_2O revealed a single dramatic difference in the imino regions of the two decamers, as shown in Table 2 and Figure 7. This region contains resonances from the Watson-Crick base-paired imino protons and is thus highly diagnostic for disruption of base-pairing. The imino region of **C_3X-10** is almost identical to that of unmodified DNA. However, in the imino region of **C_2X-10**, the proton corresponding to the crosslinked base-pair is shifted ~1 ppm upfield. This resonance is also substantially broader than the other signals and is very temperature-sensitive, disappearing at 29 °C. Even at 7 °C it shows significant exchange with bulk water (data not shown). The base-paired amino proton of C6 is also shifted upfield by 0.74 ppm relative to the control, and the G5 base-paired amino proton is shifted upfield by 0.74 ppm relative to **C_3X-10**. Taken together, these data suggest that Watson-Crick base pairing of the cross-linked bases is severely compromised in **C_2X-10**.[53,54] Despite this perturbation, the molecule retains its C_2 symmetry on the NMR timescale, as only four imino signals are observed. Were C_2 symmetry broken, eight separate resonances would have been observed. This symmetry maintenance could occur if both of the central G-C base pairs are

distorted in an identical fashion, or more likely if the molecule is in rapid equilibrium, such that any structural perturbations are averaged on the NMR time-scale. The chemical shifts are not consistent with a fully unpaired dG, but may represent an average of base-paired and non-base-paired conformers. It is significant that in the same sequence context, a two carbon disulfide crosslink between the inner cytosines in the major groove forces the duplex to become asymmetric on the NMR timescale.[24]

Finally, to clarify the structure of these decamers, we used NMR to probe their thermal denaturation, as shown in Figure 7. It is apparent that the central two base pairs of **C_3X-10** are extraordinarily stable, showing a weak proton imino signal even at 77 °C. Interestingly, **C_2X-10** appears to melt at approximately the same temperature as the control, in conflict with the circular dichroism measurements. This data can be rationalized by the fact that NMR and CD spectroscopy measure different properties of DNA structure. NMR spectroscopy examines the exchange rate of protons involved in the Watson-Crick hydrogen bonds, while CD spectroscopy measures predominantly base-stacking interactions.[48] We propose that at high temperatures, although the DNA is rapidly "breathing" and thus able to exchange imino protons with solvent, the tether constrains the molecule such that some base-stacking interactions are maintained. This explanation is consistent with the fact that clear melting transitions are not observed for the two crosslinked decamers; although the ends of the molecule are frayed, the inner bases are still stacked in approximately duplex form.

CONCLUSIONS

In light of this work and the crystal structures of DCMtase-DNA co-complexes that have been solved,[15,16] we can begin to draw some general conclusions. Both M. HaeIII and M. HhaI DCMtases methylate their substrate cytosines by everting them completely from the helical environment, thereby allowing free access of the enzyme to the substrate cytosines. The resulting "hole" in the duplex is stabilized differently in the two enzymes. In the case of M. HhaI, protein residues insert themselves into the helix and form hydrogen bonds to the orphan guanosine.[15] In contrast, the crystal structure of M. HaeIII bound to its recognition site reveals not only an extrahelical cytosine, but substantial rearrangement of the local DNA structure.[16] The most significant rearrangement occurs for the guanosine normally base-paired to the substrate cytosine, which in the structure is now weakly hydrogen-bonded to the cytosine above and across from it, displacing the guanosine that is normally its partner. This results in a gap of about eight Å between base pairs in the center of the recognition site. This dramatic rearrangement is surprising, but since the thermodynamic cost of such a distortion is not understood, it may not be as unfavorable as it appears. Indeed, since the crosslinked oligonucleotides (which are physically incapable of undergoing this type of distortion) do serve as substrates for this enzyme, it seems likely that this distortion is not obligate.

The common feature of DCMtase structures determined so far is the extrahelical cytosine. What remains unclear is exactly how much energy is required to achieve this distortion. DCMtases are unusually slow enzymes,[12] presumably due to the high kinetic barrier for disruption of the Watson-Crick base pair necessary to gain access to the substrate cytosine. Furthermore, the mode in which DCMtases induce DNA structural perturbation is unknown. Does the enzyme merely exploit the natural breathing of the helix, trapping the minute population of transiently unbase-paired cytosines? Or does it actively distort the helix to "pop" the cytosine out, and if so, from where does the energy come? Our data begin to address these questions from an experimental standpoint.

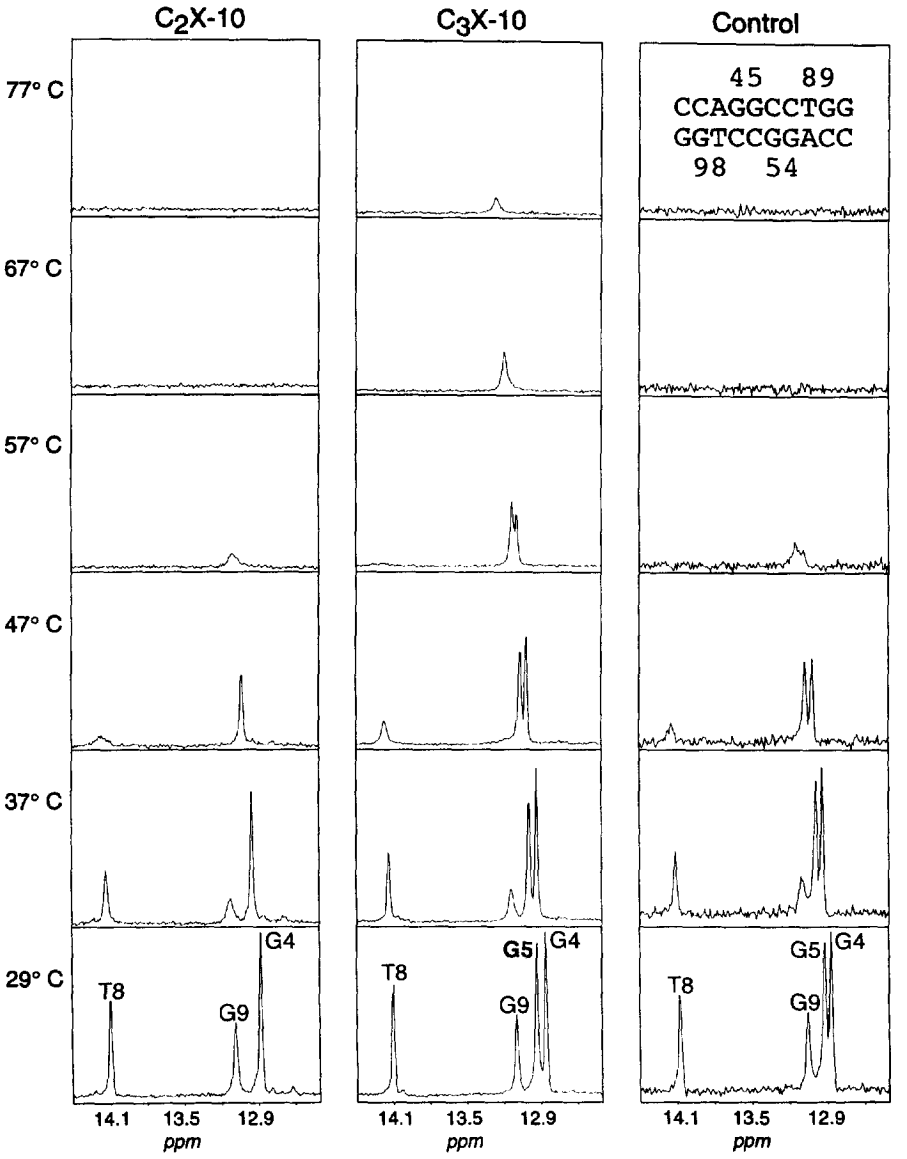


Figure 7. NMR spectra of the imino region at increasing temperatures reveals that in **C_3X-10** the central G-C base pair is stabilized while in **C_2X-10** it is destabilized.

Using disulfide crosslinking, we have manipulated the microscopic structure and dynamics of duplex DNA without affecting global structure, as seen in the NMR data. In the **C_3X** family, the central G-C base pair is stabilized relative to unmodified DNA. This stabilization, however, has little effect on the binding affinity of M. HaeIII. In the **C_2X** family, the central G-C base pair is destabilized relative to unmodified DNA. This

localized destabilization dramatically enhances the affinity of the DCMtase M. HaeIII for the oligonucleotide. This high affinity appears to be due largely to a decrease in k_{off}. The simplest explanation for the high affinity of M. HaeIII for **C_2X-18** is that the destabilized G-C base-pair causes an increase in the energy of the free oligonucleotide, thereby increasing the change in free energy for complex formation. Significantly, it has been reported that DCMtases bind tightly to sites containing mis-matched bases, or even sites lacking the substrate cytosine altogether.[55] However, this hypothesis would predict that **C_3X-18**, with a stabilized G-C base-pair, would have a decreased affinity for M. HaeIII.

This apparent conflict can be resolved if both tethers impart a net stabilization on the enzyme-oligonucleotide complex. The origin of this stabilization was alluded to earlier; in the crystal structure, a rearrangement of the active-site base-pairs occurs. This distortion does not appear to be inherently stable; the new guanosine-cytosine base-pair is twisted, and base-stacking interactions are severely compromised. Furthermore, there are few protein contacts to the guanosine opposite the substrate cytosine or to the unpaired guanosine.[16] Both tethers would prevent the "melting" of the central base-pairs seen in the crystal structure, forcing the DNA to adopt a structure similar to that seen in the M. HhaI - DNA structure. If this is in fact a more stable conformation, then we would expect that the dissociation rate for **C_2X-18** would still be lower than for **C_3X-18**, since in the former case the "product" (free DNA) is higher in energy.

EXPERIMENTAL

General. The synthesis and characterization of **C_2X-18**, **C_3X-18**, and **MTE-18** have been previously reported;[14] the shorter oligonucleotides were synthesized analogously but with DTNB activation of the thiols to form the crosslink.[30] EMSA experiments to determine equilibrium dissociation constants and equilibrium dissociation rate constants were performed as described previously.[14,39]

Reduction with 2-mercaptoethanol. ^{32}P-labeled oligonucleotides were incubated at room temperature for 13.5 hours in varying amounts of 2-mercaptoethanol in 1XTE pH 8.0 (20 µl total volume). Samples were analyzed on a 20% denaturing polyacrylamide gel.

Reduction with tris-2-carboxyethylphosphine (TCEP). ^{32}P-labeled oligonucleotides were incubated for varying lengths of time (0 to 4 h) in a buffer (20 µl total volume) containing 1.0 - 1.8 mM TCEP, 50 mM potassium acetate pH 5.0, and 100 mM NaCl. Samples were analyzed on a 20% denaturing polyacrylamide gel. The oligonucleotides were present at sub-nanomolar concentrations, so the data could be analyzed using pseudo-first-order kinetics. NMR experiments confirmed that during the reaction the TCEP did not undergo detectable oxidation.

Circular dichroism spectroscopy. Samples for CD consisted of approximately 5 µM oligonucleotide duplex in 10 mM $Na_xH_yPO_4$, 100 mM NaCl, 0.2 mM EDTA, pH 6.5. The DNA was annealed before recording spectra at 7 °C or 80 °C on a JASCO J-710 spectropolarimeter in a 1 mm pathlength cell; four separate scans were averaged for each oligonucleotide.

NMR spectroscopy general information. NMR experiments were recorded on a Bruker DMX-500 spectrometer equipped with a triple resonance probe and triple axis gradients. Measurements were made on samples between 0.75 and 1.5 mM, buffered to pH 6.5 (10 mM $Na_xH_yPO_4$, 100 mM NaCl, 0.2 mM EDTA) in either 100% D_2O or 90% H_2O and 10% D_2O at the temperatures indicated. Sequential ^{1}H assignments for the decamers were made by standard methods from ^{1}H-^{1}H double quantum filtered correlation spectroscopy (DQF-

COSY) and ^{1}H-^{1}H NOESY experiments. NMR data was processed using a Silicon Graphics Indigo 2 using FELIX (Hare Research). All proton chemical shifts are referenced to HDO (4.76 ppm). Imino ^{1}H spectra were acquired using a WATERGATE sequence to suppress the solvent peak.[56]

^{1}H-^{1}H NOESY experiments. H_2O experiments were taken in 200 μL 90:10 H_2O:D_2O in a Shigemi NMR cell. A WATERGATE pulse sequence was used for water suppression. Spectral width was set to 20 ppm, the mixing time to 50 ms, and the spectra were acquired with 256 complex points in t_1 (TPPI) and 1024 complex points in t_2. For D_2O experiments, the residual solvent peak was suppressed by presaturation. Spectral width was set to 8 ppm, the mixing time to 50 or 125 ms, with a relaxation delay of 1.5 sec and more than 175 complex points in t_1 (TPPI). 1024 complex points were collected in t_2, with 32 or more scans per FID and a relaxation delay of 1.5 sec.

^{1}H-^{1}H DQF-COSY experiments. Double quantum filtered COSY spectra were recorded in phase sensitive mode using TPPI, and presaturation of the residual solvent signal. Spectral width was set to 8 ppm. More than 175 complex points in t_1, and 1024 complex points in t_2 were collected, with 32 or more scans for each FID and a relaxation delay of 1.5 sec.

^{31}P-^{1}H hetero-TOCSY experiments. This experiment uses the pulse sequence designed by Kellogg[51,52] Data was collected on a Bruker AM-500 spectrometer. Spectra were taken in inverse mode with the sweep width in the indirect dimension set to 6 ppm (1218.59 Hz). At least 48 complex points in t_1 (TPPI), and at least 1024 complex points in t_2 were collected, with 96 or more scans for each FID. The ^{31}P 90° pulse width was calibrated using trimethyl phosphate as a standard, utilizing a BFX-5 external amplifier to adjust the power level so as to achieve a 90° pulse length that was equal to that on the ^{1}H channel. The TOCSY spin-lock time was set to about 40 ms.

REFERENCES

1. Lawley, P. D.; Brookes, P. *Biochem. J.* **1963**, *89*, 127-138.

2. Verdine, G. L. *Cell* **1994**, *76*, 197-200.

3. Nelson, H. C. M.; Bestor, T. H. *Chem. Biol.* **1996**, *3*, 419-423.

4. Kumar, S.; Cheng, X.; Klimasauskas, S.; Mi, S.; Posfai, J.; Roberts, R. J.; Wilson, G. G. *Nucl. Acids Res.* **1994**, *22*, 1-10.

5. Barlow, D. P. *Science* **1995**, *270*, 1610-1613.

6. Li, E.; Bestor, T.; Jaenisch, R. *Cell* **1992**, *69*, 915-926.

7. Volpe, P.; Iacovacci, P.; Butler, R. H.; Eremenko, T. *FEBS* **1993**, *329*, 233-237.

8. Bestor, T. H.; Coxon, A. *Curr. Biol.* **1993**, *3*, 384-386.

9. Trottier, Y.; Devys, D.; Mandel, J. L. *Curr. Biol.* **1993**, *3*, 783-786.

10. Krieg, A. M.; Yi, A. -K.; Matson, S.; Waldschmidt, T. J.; Bishop, G. A.; Teasdale, R.; Koretzky, G. A.; Klinman, D. M. *Nature* **1995**, *374*, 546-549.

11. Shen, J. -C.; Rideout III, W. M.; Jones, P. A. *Nucleic Acids Res.* **1994**, *22*, 972-976.

12. Wu, J. C.; Santi, D. V. *J. Biol. Chem.* **1987**, *262*, 4778-4786.

13. Chen, L.; MacMillan, A. M.; Verdine, G. L. *J. Am. Chem. Soc.* **1993**, *115*, 5318-5319.

14. Erlanson, D. A.; Chen, L.; Verdine, G. L. *J. Am. Chem. Soc.* **1993**, *115*, 12583-12584.

15. Klimasauskas, S.; Kumar, S.; Roberts, R. J.; Cheng, X. *Cell* **1994**, *76*, 357-369.

16. Reinisch, K. M.; Chen, L.; Verdine, G. L.; Lipscomb, W. N. *Cell* **1995**, *82*, 143-153.

17. Ferentz, A. E. and Verdine, G. L. (1994) in *Nucleic Acids and Molecular Biology* (Eckstein, F. and Lilley, D. M. J., eds) Springer-Verlag, New York

18. Lipsett, M. N. *J. Biol. Chem.* **1967**, *242*, 4067-4071.

19. Ferentz, A. E.; Verdine, G. L. *J. Am. Chem. Soc.* **1991**, *113*, 4000-4002.

20. Glick, G. D. *J. Org. Chem.* **1991**, *56*, 6746-6747.

21. Glick, G.; Osborne, S.; Knitt, D.; Marino, J. *J. Am. Chem. Soc.* **1992**, *114*, 5447-5448.

22. Ferentz, A. E.; Keating, T. A.; Verdine, G. L. *J. Am. Chem. Soc.* **1993**, *115*, 9006-9014.

23. Ferentz, A. E.; Workiewicz-Kuzera, J.; Karplus, M.; Verdine, G. L. *J. Am. Chem. Soc.* **1993**, *115*, 7569-7583.

24. Wolfe, S. A.; Verdine, G. L. *J. Am. Chem. Soc.* **1993**, *115*, 12585-12586.

25. Stevens, S.; Swanson, P.; Voss, E.; Glick, G. *J. Am. Chem. Soc.* **1993**, *115*, 1585-1586.

26. Goodwin, J. T.; Glick, G. D. *Tetrahedron Lett.* **1994**, *35*, 1647-1650.

27. Wang, H.; Osborne, S. E.; Zuiderweg, E. R. P.; Glick, G. D. *J. Am. Chem. Soc.* **1994**, *116*, 5021-5022.

28. Wang, H.; Zuiderweg, E. R. P.; Glick, G. D. *J. Am. Chem. Soc.* **1995**, *117*, 2981-2991.

29. Allerson, C. R.; Verdine, G. L. *Chem. Biol.* **1995**, *2*, 667-682.

30. Wolfe, S. A.; Ferentz, A. E.; Grantcharova, V.; Churchill, M. E. A.; Verdine, G. L. *Chem. Biol.* **1995**, *2*, 213-221.

31. Sigurdsson, S. T.; Tuschl, T.; Eckstein, F. *RNA* **1995**, *1*, 575-583.

32. Chaudhuri, N. C.; Kool, E. T. *J. Am. Chem. Soc.* **1995**, *117*, 10434-10442.

33. Osborne, S. E.; Völker, J.; Stevens, S. Y.; Breslauer, K. J.; Glick, G. D. *J. Am. Chem. Soc.* **1996**, *118*, 11993-12003.

34. Osborne, S. E.; Ellington, A. D. *BioMed. Chem. Lett.* **1996**, *6*, 2339-2342.

35. Goodwin, J. T.; Osborne, S. E.; Scholle, E. J.; Glick, G. D. *J. Am. Chem. Soc.* **1996**, *118*, 5207-5215.

36. Osborne, S. E.; Cain, R. J.; Glick, G. D. *J. Am. Chem. Soc.* **1997**, *119*, 1171-1182.

37. MacMillan, A. M.; Verdine, G. L. *J. Org. Chem.* **1990**, *55*, 5931-5933.

38. Riddles, P. W.; Blakeley, R. L.; Zerner, B. *Methods Enzymol.* **1983**, *91*, 49-60.

39. Chen, L. (1994) *Ph.D. Thesis*, Harvard University,

40. Lesser, D. R.; Kurpiewski, M. R.; Jen-Jacobson, L. *Science* **1990**, *250*, 776-786.

41. Houk, J.; Whitesides, G. M. *J. Am. Chem. Soc.* **1987**, *109*, 6825-6836.

42. Falcomer, C. M.; Meinwald, Y. C.; Choudhary, I.; Talluri, S.; Milburn, P. J.; Clardy, J.; Scheraga, H. A. *J. Am. Chem. Soc.* **1992**, *114*, 4036-4042.

43. Gilbert, H. F. *Methods Enzymol.* **1995**, *251*, 8-30.

44. Podlaha, J.; Podlahová, J. *Collection Czechoslov. Chem. Commun.* **1973**, *38*, 1730-1736.

45. Burns, J. A.; Butler, J. C.; Moran, J.; Whitesides, G. M. *J. Org. Chem.* **1991**, *56*, 2648-2650.

46. Heinemann, U.; Alings, C. *J. Mol. Biol.* **1989**, *210*, 369-381.

47. Marky, L. A.; Breslauer, K. J. *Biopolymers* **1987**, *26*, 1601-1620.

48. Cantor, C. R.; Warshaw, M. M. *Biopolymers* **1970**, *9*, 1059-1077.

49. Saenger, W. (1984) *Principles of Nucleic Acid Structure*, Springer-Verlag, New York

50. Hare, D. R.; Wemmer, D. E.; Chou, S. -H.; Drobny, G.; Reid, B. R. *J. Mol. Biol.* **1983**, *171*, 319-336.

51. Kellogg, G. W. *J. Mag. Res.* **1992**, *98*, 176-182.

52. Kellogg, G. W.; Szewczak, A. A.; Moore, P. B. *J. Am. Chem. Soc.* **1992**, *114*, 2727-2728.

53. Woodson, S. A.; Crothers, D. M. *Biochemistry* **1988**, *27*, 436-445.

54. Woodson, S. A.; Crothers, D. M. *Biochemistry* **1988**, *27*, 3130-3141.

55. Klimasauskas, S.; Roberts, R. J. *Nucleic Acids Res.* **1995**, *23*, 1388-1395.

56. Piotto, M.; Saudek, V.; Sklenár, V. *J. Biomol. NMR* **1992**, *2*, 661-665.

ACKNOWLEDGMENTS

We thank Mary T. Didiuk and Monya L. Baker for a careful reading of this manuscript. This work was supported by the National Institute of Health (GM 44853).

(*Received* 15 *April* 1997; *accepted* 30 *June* 1997)