Optimization of Alternate-Strand Triple Helix Formation at the 5′CpG3′ and 5′GpC3′ Junction Steps

Christophe Marchand,[‡] Jian-Sheng Sun,*[,‡] Christian Bailly,[§,||] Michael J. Waring,[§] Thérèse Garestier,[‡] and Claude Hélène[‡]

Laboratoire de Biophysique, Muséum National d'Histoire Naturelle, INSERM U201, CNRS URA481, 43 rue Cuvier, 75231 Paris Cedex 05, France, and Department of Pharmacology, University of Cambridge, Tennis Court Road, Cambridge CB2 1QJ, U.K.

Received March 17, 1998; Revised Manuscript Received June 23, 1998

ABSTRACT: Oligonucleotide-directed triple helix formation normally requires a long tract of oligopyrimidine·oligopurine sequence. This limitation can be partially overcome by alternate-strand triple helix (or switch triple helix) formation which enables recognition of alternating oligopurine/oligopyrimidine sequences. The present work is devoted to the optimization of switch triple helix formation at the 5′CpG3′ and 5′GpC3′ junction steps by combination of base triplets in Hoogsteen and in reverse Hoogsteen configurations. Rational design by molecular mechanics was first carried out to study the geometrical constraints at different junction steps and to propose a "switch code" which would optimize the interactions at junctions. These predictions were further checked and validated experimentally by gel retardation and DNase I footprinting assays. It was shown that the choice of an appropriate linker nucleotide in the switching third strand plays an important role in the interaction between oligonucleotides and alternating oligopurine/oligopyrimidine target sequences at different junctions: (i) the addition of a cytosine at the junction level in the oligonucleotide optimizes the crossover at the 5′CpG3′ junction, whereas (ii) the best crossover at the 5′GpC3′ junction step is achieved without any additional nucleotide. These results provide a useful guideline to extend double-stranded DNA sequence recognition by switch triple helix formation.

Oligonucleotides can bind to the major groove of double-stranded DNA containing oligopyrimidine·oligopurine sequences in a sequence-specific manner (*1, 2*). The resulting triple helix is able to interfere selectively with gene expression. This approach provides a rational basis for the development of new tools in molecular biology and for therapeutic applications.

In this so-called anti-gene strategy, it is well established that an oligonucleotide containing thymines and cytosines can form a triple helix where the third strand adopts a parallel orientation with respect to the oligopurine strand of the targeted DNA. The T and protonated C residues of the third strand are involved in Hoogsteen base pairing and form the two canonical base triplets T·AxT and C·GxC+.

On the other hand, (G,A)- and (G,T)-containing third strands allow triple helix formation at neutral pH. In a (G,A)-motif triplex (*3*−*8*), the oligopurine strand recognizes a double-stranded DNA target in an antiparallel orientation with respect to the oligopurine sequence. In a (G,T)-motif triplex (*9*−*16*), an oligonucleotide containing thymines and guanines binds in an antiparallel or a parallel orientation with respect to the purine target sequence, depending on the sequence of the third strand. Its orientation changes from parallel to antiparallel when the number of 5′GpT3′ and 5′TpG3′ steps increases and the length of G and T tracts decreases (*12, 16*).

There is therefore a sequence restriction in the "anti-gene" strategy, since the DNA target sequence must possess all purines on the same strand. However, recent studies have suggested the possibility of extending the range of DNA targets for triple helix formation in two ways. (i) When the oligopyrimidine·oligopurine sequence is interrupted by a single or double base pair inversion, triple helix formation can still occur by using a natural or modified nucleotide facing the inverted purine·pyrimidine base pair(s) (*17*−*23*) by dimerization of adjacent triplex-forming oligonucleotides (*24, 25*), or by using an intercalator-containing oligonucleotide in which the intercalator is internally attached at the appropriate place where it can intercalate at the inverted base pair position (*26*). (ii) When the target sequence is composed of alternating oligopurine/oligopyrimidine tracts, an oligonucleotide containing several appropriately linked triplex-forming domains can engage in alternate-strand triple helix (or switch triple helix) formation. Such an oligonucleotide zigzags within the major groove, switching from one oligopurine strand to another at the 5′RpY3′ and 5′YpR3′ junction steps. Utilizing this principle, two distinct approaches have been developed: (i) when the adjacent triple helices are formed by motifs involving the same hydrogen

* To whom correspondence should be addressed at the Laboratoire de Biophysique, 43 rue Cuvier, 75231 Paris Cedex 05, France. E-mail: sun@mnhn.fr. Fax: 33-1-40793705.

[‡] Muséum National d'Histoire Naturelle.

[§] University of Cambridge.

[||] Present address: INSERM U124, Centre Oscar Lambret, IRCL, Place de Verdun, 59045 Lille, France.

S0006-2960(98)00618-7 CCC: $15.00
Published on Web 09/01/1998

bonding configuration (Hoogsteen or reverse Hoogsteen), a suitable linker must be used to join the two 3′- or 5′-ends of the third strands via their phosphodiester backbones (*27*−*31*) or terminal bases (*32*); (ii) when they involve motifs switching from Hoogsteen to reverse Hoogsteen hydrogen-bonding (or vice versa) (*16*, *33*−*38*), the juxtaposed portions of the third strand could be linked through a natural 5′−3′ phosphodiester bond.

The present work is devoted to the amelioration of "switch triple helix" formation by establishing rules to optimize the oligonucleotide sequence so that it can cross the major groove at two types of junctions, namely, 5′CpG3′ and 5′GpC3′ steps. The predictions by molecular modeling have been checked and validated by gel retardation and DNase I footprint assays.

MATERIALS AND METHODS

Molecular modeling by conformational energy-minimization was carried out using JUMNA (version 7) program package (*39*). Neither water nor positively charged counterions were explicitly included in the energy minimization. However, their effects were simulated by a sigmoidal, distance-dependent, dielectric function (*40*) and by assignment of a half negative charge for each phosphate group. Computations were carried out on a Silicon Graphics 4D/420GTXB dual processor workstation.

The coordinates of triple helices were derived from the previously published B-like triple helix (*41*, *42*) which is now widely supported by many NMR and vibrational spectroscopic studies. Typically, the triple helices were 10 base triplets in length with the junction step occurring at the center. The three last base triplets at both ends were restrained to a mononucleotide symmetry in order to decrease the end effects and to focus on the effect at the junction (the central four base triplets). A manual docking was sometimes required around the junction in order to establish an appropriate interaction and to avoid strong steric clashes during initial steps of minimization. The total complexation energy (E_{TOT}) was decomposed in terms of intermolecular interactions (E_{DH-III}) between the third strand (III) and the target double-helical DNA (DH), as well as the conformational deformation energy of the double helix (ΔE_{DH}) and the third strand (ΔE_{III}). Conformational deformation energies were evaluated as the difference between the corresponding energetic components before and after complexation of the third strand. It should be noted that this evaluation is approximate, especially for the third strand since the conformation of a free single-stranded DNA is generally less well-defined than a free double-stranded DNA. However, these approximations were necessary to take into account the effect of base composition.

Chemicals. Deoxyribonuclease I (DNase I) was purchased from Sigma Chemicals Co. The restriction enzymes *Eco*RI, *Bam*HI, *Hin*dIII, *Pvu*II, and polynucleotide kinase were obtained from Biolabs. AMV reverse transcriptase was purchased from Eurogentec. Nucleoside triphosphates labeled with ^{32}P (γ-ATP and α-dATP) were from ICN.

Oligonucleotides were synthesized on an automatic synthesizer by Eurogentec using phosphoramidite chemistry. Full-length oligonucleotides were purified by polyacrylamide gel electrophoresis under denaturing conditions. Oligonucleotide concentrations were determined by the UV absorption at 260 nm on a Kontron Uvikon 860 Spectrophotometer at 20 °C using the nearest-neighbor method for calculating molar extinction coefficients (*43*). The 54 bp duplexes used in this study were labeled at the 5′-end using [γ-^{32}P]ATP and polynucleotide kinase.

DNA Purification and Labeling. To clone the 28 bp alternating oligopyrimidine·oligopurine sequences containing either a 5′CpG3′ or a 5′GpC3′ step, double-stranded oligonucleotides were designed to be inserted between the *Hin*dIII and *Bam*HI sites of the pUC12 plasmid. The cloning was carried out as previously described (*44*) with the following modifications. The plasmid was cut by *Eco*RI and 3′-^{32}P-end-labeled using [α-^{32}P]dATP (6000 Ci/mmol) and AMV reverse transcriptase. The 148 bp *Eco*RI−*Pvu*II DNA fragment for footprinting was prepared by a final *Pvu*II digestion. The digestion products were separated on an 8% polyacrylamide gel under nondenaturing conditions in TBE buffer (89 mM Tris−borate, pH 8.3, 1 mM EDTA).

Gel Retardation Assay. Double-stranded target DNA radiolabeled at one 5′-end (20 nM) was incubated in the presence of the third strand in 10 μL of a solution containing 50 mM HEPES, pH 7.2, 50 mM NaCl, 10 mM $MgCl_2$, 5 μg of tRNA, and 20% sucrose. After overnight incubation at a controlled temperature, samples were loaded on a 12% nondenaturing polyacrylamide gel. Electrophoresis was performed in 50 mM HEPES, pH 7.2, and 10 mM $MgCl_2$. After drying, the gel was exposed to a phosphorimager storage screen, and quantification was performed on a Phosphorimager densitometer (Molecular Dynamics). The uncertainty of quantification was estimated to about 10% by repeated experiments.

DNase I Footprinting. Experiments were carried out as previously described (*44*) with the following modifications. Samples (2 μL) of the labeled DNA fragment were incubated with 1 μL of a 10× concentrated solution of the third strand oligonucleotide. All reactions were performed in 10 mM Tris-HCl, pH 7.2, 50 mM NaCl, and 10 mM $MgCl_2$ in the presence of 0.2 mM spermine. To ensure equilibrium of the binding reaction, samples were incubated overnight at 37 °C.

RESULTS

Prediction by Molecular Modeling. Molecular modeling shows that the geometry of two-third strands which bind to oligopurine sequences on the alternate strands of a target duplex at the 5′CpG3′ junction step (5′YpR3′) (Figure 1, top part) is distinctly different from that which pertains at the 5′GpC3′ junction step (5′RpY3′) (Figure 2, top part). Different approaches were used to cope with the conformational requirements for designing appropriate oligonucleotides to switch at these junctions. When the two triplexes are formed by combination of the C·GxG base triplets in Hoogsteen and reverse Hoogsteen configuration, the modeling studies show that the 5′- and 3′-ends of two third strands can easily be joined by using a single phosphodiester unit at the 5′GpC3′ junction (Figure 2, bottom part), whereas the 5′- and 3′-ends of two third strands are too far away at the 5′CpG3′ junction. Therefore, additional nucleotide(s) should be required to facilitate the switching over of the third strand at the 5′CpG3′ junction (Figure 1, bottom part). Molecular

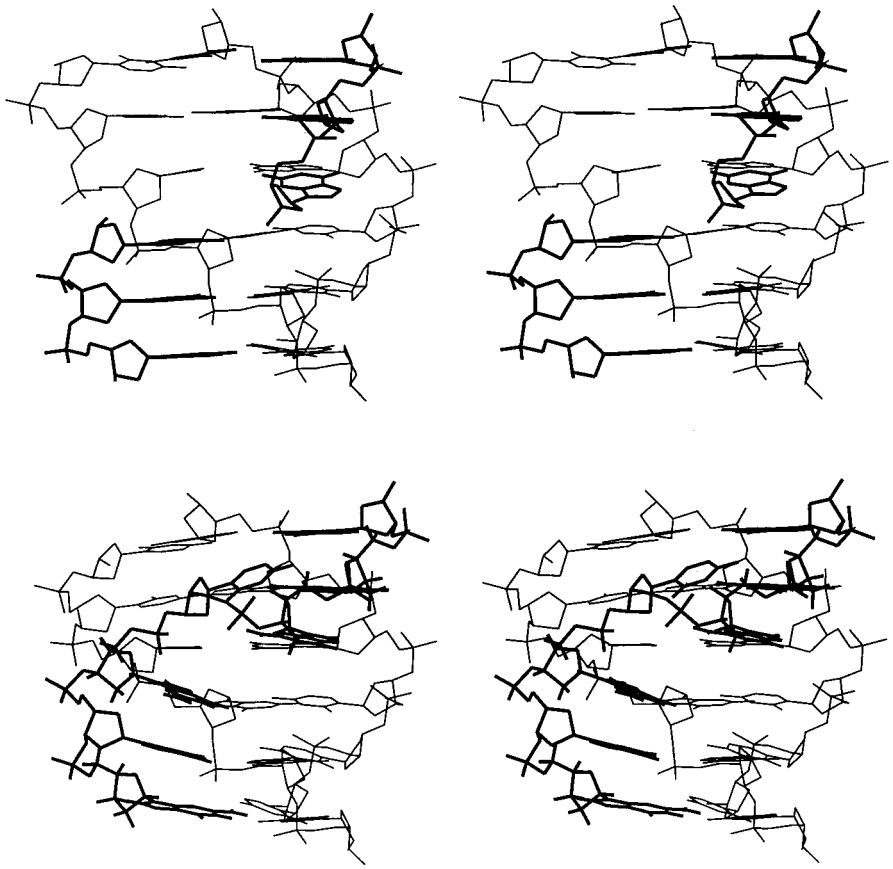

FIGURE 1: Stereoview of energy-minimized models of triplexes in which the underlying target duplexes have alternating oligopyrimidine/oligopurine sequence containing a 5′CpG3′ junction step. Three base triplets are shown on each side of the junction step. The third strands are indicated in boldface lines. Hydrogen atoms are not drawn for clarity. Top: The two third strands are not linked together. Bottom: a cytosine was added in the third strand to facilitate the switching over of the third strand (CG+C in Table 1, see text for details).

modeling by energy minimization was used to devise an appropriate linkage for bridging the gap between two third strands in order to stabilize the junction and to increase the specificity by establishing hydrogen bonding when possible.

5′CpG3′ Junction Step. In Table 1, a combination of two C·GxG base triplets has been considered. The first triplet is in the reverse Hoogsteen-like configuration which occurs in either an antiparallel (G,T)- or a (G,A)-motif triple helix on the left half of the target sequence, while the second triplet is in Hoogsteen-like configuration, required for a parallel (T,G)-motif triple helix on the right half of the target sequence.

The triplexes (from CG+A to CG+C, see Table 1) contain an extra nucleotide (A, G, U, T, and C, respectively) in the third strand at the 5′CpG3′ junction step. All of them have lower complexation energies than that of the triplex CG which contains no additional nucleotide. These improved complexation energies are due mainly to strong intermolecular interactions between the third strand and the target double-helical DNA. It can be seen that both the nature and the number of the extra nucleotide(s) play important roles. As a matter of fact, the introduction of a single cytosine into the third strand at the junction step serving as a linker turns out to be the most energetically preferred switch triplex (triplex CG+C) compared to the triplexes containing any of the other nucleotides (A,G,T). It is also worth noting that the addition of two cytidines produces a less favored triplex (CG+CC).

The model of the best switch triplex (CG+C) at the 5′CpG3′ junction step (Figure 1, bottom part) indicates that the phosphodiester backbone of the extra cytosine in the third strand effectively acts as an appropriate linker bridging the gap between the Hoogsteen and reverse Hoogsteen parts of the switch triple helix. In addition, this cytosine engages in three hydrogen bonds with the second neighbor C·GxG base triplet in reverse Hoogsteen-like configuration. Therefore, it not only provides the appropriate spacer for linking two parts of the third-strand oligonucleotide, but also confers specific recognition of the 5′CpG3′ junction step, provided the reverse Hoogsteen target sequence contains adjacent G's at the junction.

5′GpC3′ Junction Step. In Table 2, the triplexes are composed of two combined C·GxG base triplets, one in Hoogsteen-like configuration and the other in reverse Hoogsteen-like configuration. In each case, the first C·GxG base triplet is involved in a parallel (T,G)-motif triple helix on the left half of the target sequence, while the second C·GxG base triplet can be in either an antiparallel (G,T)- or a (G,A)-motif triple helix on the right half of the target sequence.

When one guanosine in the third strand was removed at the 5′GpC3′ junction step, either from the reverse Hoogsteen or from the Hoogsteen-like portion as in the triplexes GC-G1 and GC-G2 (Table 2), respectively, the complexation energy (E_{TOT}) proved lower than that of the full-length triplex GC1 which was obtained by directly attaching two C·GxG triplets in different hydrogen bonding configuration at the

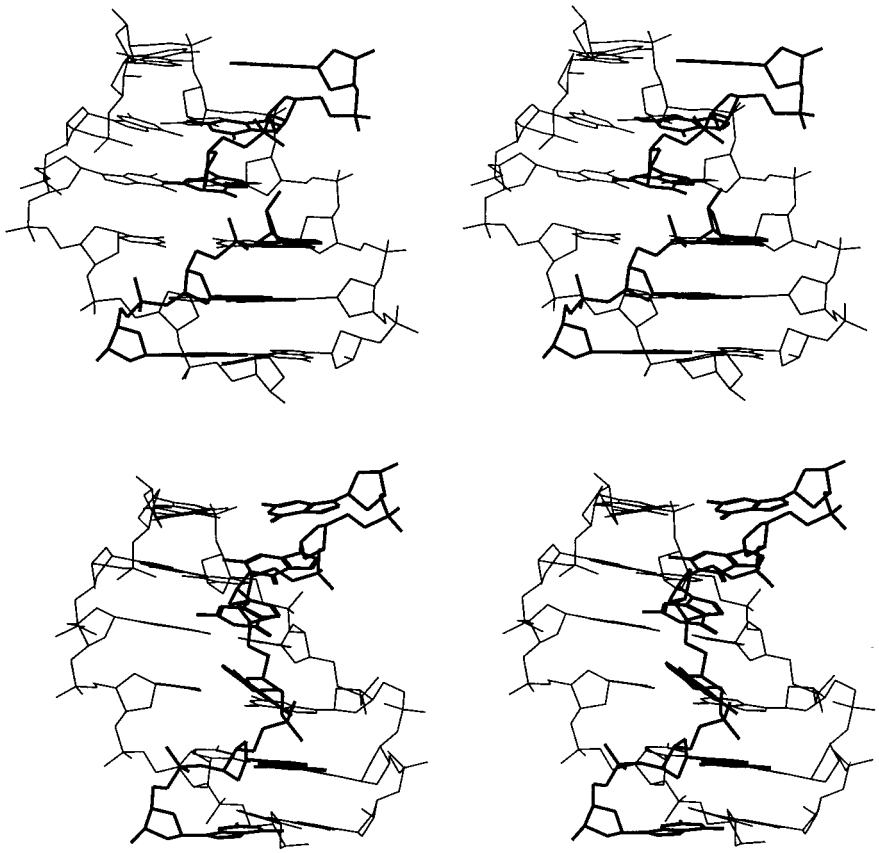

FIGURE 2: Stereoview of energy-minimized models of triplexes in which the underlying target duplexes have alternating oligopurine/oligopyrimidine sequence containing a 5′GpC3′ junction step. Three base triplets are shown on the each side of the junction step. The third strands are indicated in boldface lines. Hydrogen atoms are not drawn for clarity. Top: The two third strands are not linked together. Bottom: The two third strands are linked together by a natural 5′−3′ phosphodiester linkage (GC2 in Table 2, see text for details).

Table 1: Analysis of the Complexation Energies of the Energy-Minimized Switched Triple Helices at the 5′GpC3′ Junction Step (See the Text for Details)[a]

$^{5'}$CpG$^{3'}$ junction:		3′GGGGGCCCCC5′ 5′CCCCCGGGGG3′			
triplex	strand III	ΔE_{DH}	E_{DH-III}	ΔE_{III}	E_{TOT}
CG	$^{5'}$GGGGGGGGGG$^{3'}$	+43	−284	+76	−165
CG+A	$^{5'}$GGGGG*A*GGGGG$^{3'}$	+67	−333	+69	−197
CG+G	$^{5'}$GGGGG*G*GGGGG$^{3'}$	+64	−322	+85	−173
CG+U	$^{5'}$GGGGG*U*GGGGG$^{3'}$	+52	−301	+51	−198
CG+T	$^{5'}$GGGGG*T*GGGGG$^{3'}$	+54	−299	+58	−187
CG+C	$^{5'}$GGGGG*C*GGGGG$^{3'}$	+55	−318	+54	−209
CG+CC	$^{5'}$GGGGG*CC*GGGGG$^{3'}$	+59	−303	+53	−191

[a] The underlined letters show the nucleotides interacting with the purine strand in the target sequence through reverse Hoogsteen-like hydrogen bonds. The italic letters indicate that those nucleotides are primarily used as linkers to bridge the gap between Hoogsteen and reverse Hoogsteen-like parts of the third strand. Energies are given in kilocalories per mole (1 kcal = 4.18 kJ).

Table 2. Analysis of the Complexation Energies of the Energy-Minimized Switched Triple Helices at the 5′GpC3′ Junction Step (See the Text for Details)[a]

$^{5'}$GpC$^{3'}$ junction:		3′CCCCCGGGGG5′ 5′GGGGGCCCCC3′			
triplex	strand III	ΔE_{DH}	E_{DH-III}	ΔE_{III}	E_{TOT}
GC1	$^{5'}$GGGGGGGGGG$^{3'}$	+58	−268	+83	−127
GC2	$^{5'}$GGGGG**G**GGGG$^{3'}$	+50	−263	+72	−141
GC-G1	$^{5'}$GGGGG-GGGG$^{3'}$	+47	−245	+66	−132
GC-G2	$^{5'}$GGGG-GGGGG$^{3'}$	+48	−246	+69	−129

[a] The underlined letters show the nucleotides interacting with the purine strand in the target sequence through reverse Hoogsteen-like hydrogen bonds. The hyphen symbolizes a standard phosphodiester unit. The boldface letter shows that the nucleotide adopts a syn glycosidic conformation. Energies are given in kilocalories per mole (1 kcal = 4.18 kJ).

junction. These two energetically favored triplexes appear to profit from a greater extent of energy reduction in terms of deformation energies in both the double-helical DNA (ΔE_{DH}) and the third strand (ΔE_{III}) than a gain in energy of intermolecular interaction (E_{DH-III}) between the target DNA and the third strand. However, it turns out that another alternative, namely, full-length triplex GC2 in which the first guanine involved in the reverse Hoogsteen part of the third strand is in syn glycosidic conformation, appears to be the most energetically preferred. Thus, this C·GxG base triplet actually establishes hydrogen bonds in Hoogsteen-like configuration. By doing so, it affords a significant reduction of deformation energies (ΔE_{DH} and ΔE_{III}) in parallel with a limited loss in intermolecular interaction (E_{DH-III}) as compared to that of the corresponding triplex GC1 (Table 2). The model of the triplex GC2 shows that the flipping of guanosine from anti to syn glycosidic conformation serves to preserve intra-third-strand base stacking at the 5′GpC3′ junction (Figure 1, bottom part).

Experimental Investigation. Two 54 base pair duplexes were synthesized to study switch triplex formation at the 5′CpG3′ and 5′GpC3′ steps. These target molecules are composed of two adjacent sequences (box I and box II) in which the two oligopurine tracts alternate across the two strands. Box I contains a 14 bp oligopurine·oligopyrimidine sequence with two 5′GpA3′/5′ApG3′ steps and a long T

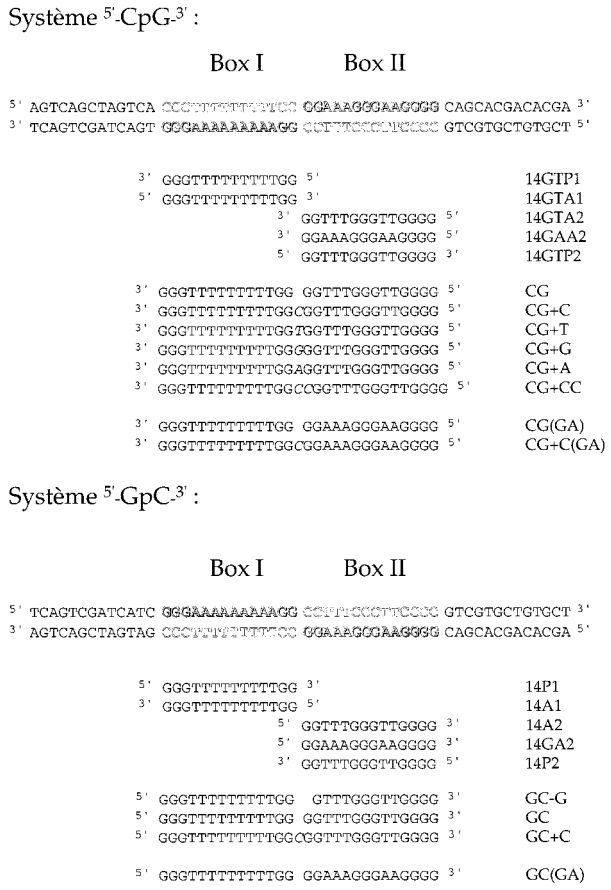

FIGURE 3: Sequences of oligonucleotides used for optimizing switch triple helix formation at the 5′CpG3′ and 5′GpC3′ junction steps.

stretch. According to previously established rules (*12*), a (T,G)-containing oligonucleotide (14GTP1 and 14P1) should bind in a parallel orientation with respect to the oligopurine strand in box I. On the other hand, box II contains a 14 bp oligopyrimidine·oligopurine sequence with four 5′GpA3′/5′ApG3′ steps; thus, a (G,T)- or (G,A)-containing oligonucleotide (14GTA2/14GA2, and 14A2/14GA2) should bind in antiparallel orientation with respect to the oligopurine strand in box II. Four further oligonucleotides, 14GTA1, 14GTP2, 14A1, and 14P2, were synthesized having the opposite orientation with respect to 14GTP1/14P1 and 14GTA2/14GA2/14A2/14GA2, respectively, and were used as controls for both systems.

Gel retardation and DNase I footprinting experiments were carried out to assess the binding of the third strands, instead of commonly employed thermal DNA denaturation by UV absorption spectroscopy because of weak absorbance changes upon binding of (G,T)- or (G,A)-containing third strand to the target sequence. The preliminary gel retardation experiments performed at 4 °C on the radiolabeled 54 bp duplexes with 20 μM 14-mers showed that only the oligonucleotides 14GTA2 and 14A2 were able to exhibit a band shift of weak intensity (data not shown). This retardation disappeared when the gel was run at room temperature.

Switch Triplex Formation at the 5′CpG3′ Junction. A series of switch oligonucleotides designed according to the principles outlined above were tested for their ability to bind to the 28 bp alternating oligopyrimidine/oligopurine target sequence at the 5′CpG3′ junction (Figure 3, top panel). The 28-mer oligonucleotide CG containing only G and T corresponds to the direct linkage of 14GTP1 to 14GTA2 through a natural 3′−5′ phosphodiester bond. As molecular modeling suggested that an additional base in the middle of the third strand should facilitate the crossover at the 5′CpG3′ junction, four 29-mer oligonucleotides derived from CG in which an A, G, T, or C was inserted as a linker (CG+A, CG+G, CG+T, and CG+C, respectively) were synthesized. A 30-mer oligonucleotide CG+CC containing two cytosines as a linker was also synthesized. As a (G,A)-containing oligonucleotide (14GA2) can be used to target box II in a reverse Hoogsteen configuration, two (G,T,A)-containing oligomers corresponding to the linkage of 14GTP1 to 14GA2 without or with an additional cytosine [CG(GA) and CG+C(GA), respectively] were also synthesized.

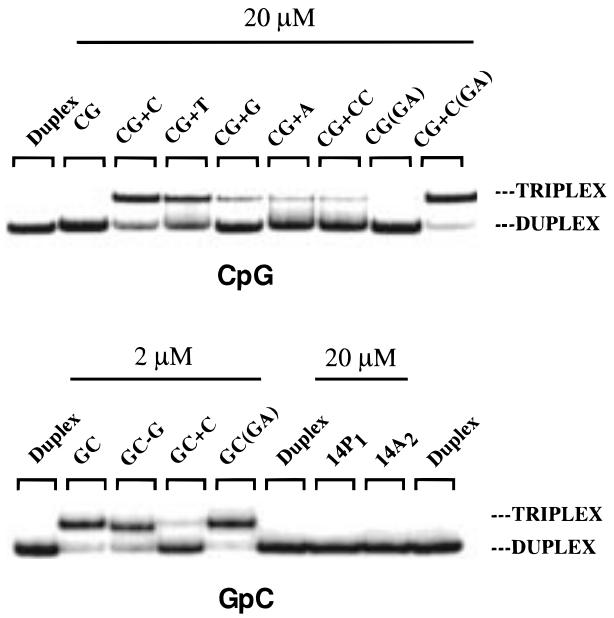

FIGURE 4: Gel retardation assays for the binding of switching oligonucleotides to the alternating oligopurine/oligopyrimidine double-helical sequences containing a 5′CpG3′ (top panel) or 5′GpC3′ (bottom panel) junction step. Experiments were carried out at pH 7.2 and 20 °C.

Gel retardation experiments carried out at 20 °C with 20 μM switch oligonucleotides (Figure 4, top panel) evidenced a strong band shift for the oligomers CG+C and CG+C(GA) as compared to their controls, CG and CG(GA), respectively, for which no band shift was detected. The retarded band migration is ascribed to triplex formation. A significant band shift was also observed with the oligomer CG+T. Other oligomers (CG+A, CG+G, and CG+CC) exhibited a weakly retarded band. The ranking according to the relative binding of these switch oligonucleotides to the alternating oligopurine/oligopyrimidine target sequences at the 5′CpG3′ junction is as follows: CG+C(GA) ∼ CG+C > CG+T > CG+G > CG+A > CG+CC ≫ CG ∼ CG-(GA). Based on this first round of screening, the dissociation constants (K_D) were determined by gel retardation experiments for the most promising 29-mer switch oligonucleotides, CG+C and CG+C(GA), and their 28-mer control oligomers, CG and CG(GA), respectively (Table 3). It can be seen that the newly designed switch oligonucleotides bind better to the alternating oligopyrimidine/oligopurine target sequence containing a 5′CpG3′ junction step by more than 1 order of magnitude.

Table 3: Dissociation Binding Constants (K_D) Determined by Gel Retardation Assays at 20 °C of the Switch Oligonucleotides Targeted to the Alternating Oligopurine/Oligopyrimidine Sequences Containing either a 5′CpG3′ or a 5′GpC3′ Junction Step (See Materials and Methods for Details)[a]

junction	switch oligonucleotide	K_d (μM)
5′CpG3′	CG	>20
	CG+C	2
	CG(GA)	>20
	CG+C(GA)	2
5′GpC3′	GC-G	1.2
	GC	0.6
	GC+C	10
	GC(GA)	0.4

[a] The uncertainity of the K_D value is estimated to about 10%.

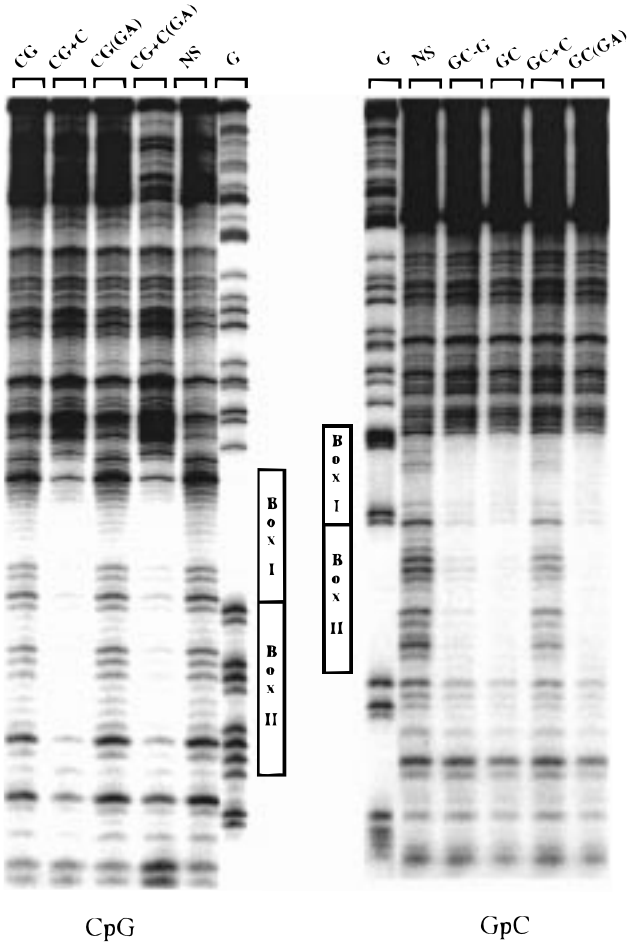

FIGURE 5: DNase I footprinting experiments of switching oligonucleotides bound to the alternating oligopurine/oligopyrimidine double-helical sequences containing a 5′CpG3′ (left panel) or 5′GpC3′ (right panel) junction step. Experiments were carried out at pH 7 and 37 °C.

DNase I footprinting experiments performed at 37 °C on the 148 bp *Eco*RI−*Pvu*II restriction fragment containing a 28 bp alternating oligopyrimidine/oligopurine target sequence (Figure 5, left panel) showed that both the 29-mers CG+C and CG+C(GA) were able to produce a clear footprint on both boxes at 20 μM. Under the same conditions, the 28-mers CG1 and CG1(GA) failed to show any notable footprint. These results are fully consistent with those obtained by gel retardation experiments.

Switch Triplex Formation at the 5′GpC3′ Junction. A number of designed switch oligonucleotides were tested for their ability to bind to the 28 bp alternating oligopurine/oligopyrimidine target sequence at the 5′GpC3′ junction (Figure 3, bottom panel). The 28-mer oligonucleotide GC containing only G and T corresponds to the direct linkage of 14P1 to 14A2 through a natural 3′−5′ phosphodiester bond. A 27-mer GC-G was also investigated because the molecular modeling suggested that a guanine could be removed from the third-strand oligonucleotide at the 5′GpC3′ junction. A 29-mer GC+C derived from the 28-mer GC in which a cytosine has been inserted between the two linked 14-mers was used as a control. This oligonucleotide gave the highest stability at the 5′CpG3′ junction (see above). Another 28-mer (G,T,A)-containing oligomer, GC(GA), which corresponds to the linkage of 14P1 to 14GA2 through a natural 3′−5′ phosphodiester bond, was also synthesized as an alternative to the oligomer GC.

Gel retardation experiments carried out at 20 °C with 2 μM switch oligonucleotides (Figure 4, bottom panel) revealed a strong band shift for the oligomers GC, GC-G, and GC-(GA), indicative of triplex formation. As expected, the control oligomer GC+C showed only a very weakly shifted band. It should be noted that none of the 14-mers succeeded in producing any band retardation even at 10-fold higher concentration (20 μM) than the switch oligonucleotides under the same conditions. The ranking according to the relative binding of these switch oligonucleotides to an alternating oligopurine/oligopyrimidine target sequence containing a 5′GpC3′ junction step is as follows: GC(GA) ∼ GC > GC-G ≫ GC+C. Dissociation constants (K_D) were determined by gel retardation experiments for these oligonucleotides (Table 3). It can be seen that the binding of the 28-mers GC and GC(GA) to the alternating oligopurine/oligopyrimidine target containing a 5′GpC3′ junction step is about twice as strong as for the 27-mer GC-G, whereas the addition of an extra cytosine (triplex GC+C) drops the binding by a factor of 20.

DNase I footprinting experiments performed at 37 °C on the 148 bp *Eco*RI−*Pvu*II restriction fragment containing the 28 bp alternating oligopurine/oligopyrimidine target sequence in the presence of 20 μM oligonucleotides revealed a strong footprint for the 28-mers [GC and GC(GA)] and the 27-mer GC-G, but not for the 29-mer GC+C (Figure 5, right panel). However, the footprint observed with the oligomer GC-G was slightly less pronounced than that of the full-length oligomers GC and GC(GA). These results are again in good agreement with the gel retardation experiments.

DISCUSSION AND CONCLUSION

Molecular modeling showed that the structures of switch triple helices containing a 5′CpG3′ junction are distinctly different from those containing a 5′GpC3′ junction. This distinction applies to all 5′YpR3′ and 5′RpY3′ junction steps. In general, the addition of one (or two) nucleotide is required to facilitate the switching over the 5′YpR3′ junction. The number and nature of additional nucleotide(s) depend on the combination of base triplets involved in switch triplex formation. In contrast, the removal of one nucleotide facing the junction step in the switch oligonucleotide could facilitate its crossover at the 5′RpY3′ junction. This is due to

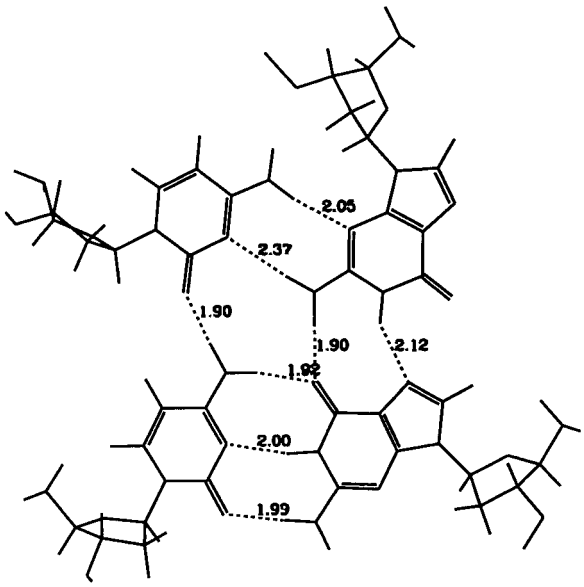

FIGURE 6: Base quadruplet involving the added cytosine and the reverse Hoogsteen C·GxG base triplet as depicted in the triplex CG+C (see Figure 1).

decreased unfavorable steric contacts and a concomitant increase in the base stacking interaction. However, whether removing one nucleotide will enhance binding depends again upon the combination of base triplets involved in switch triplex formation (*45*). The present studies were focused on the 5′CpG3′ and 5′GpC3′ junction steps with C·GxG base triplets on the both Hoogsteen and reverse Hoogsteen sides of the junction.

5′CpG3′ Junction. The experimental data are consistent with the predictions of molecular modeling. The best switch oligonucleotide for recognizing an alternating oligopyrimidine/oligopurine sequence containing a 5′CpG3′ junction by alternate-strand triple helix formation is constructed by adding one cytosine in the middle. This originates from two contributions: (i) the sugar−phosphate moiety of the added cytidine serves as a linker sufficiently long to permit strand crossover by reducing the constraints on the phosphodiester backbone and the subsequent backbone distortion at the junction; (ii) the cytosine moiety makes an unusual C·G*G-C base quadruplet (Figure 6). This unusual quadruplet is also involved in strand crossover at the 5′TpG3′ junction step, and has been evidenced by chemical probing (de Bizemont et al., to be published).

5′GpC3′ Junction. Using the same approach, it has been shown that the prediction of molecular modeling for best recognition of an alternating oligopurine/oligopyrimidine sequence containing a 5′GpC3′ junction step by alternate-strand triple helix formation is upheld by the experimental results. The optimized switch oligonucleotide is constructed by making a natural 3′−5′ phosphodiester backbone in the middle of the oligonucleotide at the junction level. As is shown by the gel retardation experiments, the removal of a guanine at the junction (GC-G) in the third strand leads to decreased triplex stability, but is less penalizing than an additional cytosine in the middle of the oligonucleotide (GC+C). This is because the strand crossover at the 5′GpC3′ junction step, proceeding from the last C·GxG base triplet in the Hoogsteen configuration to the first C·GxG triplet in the reverse Hoogsteen configuration, can be well accommodated without too severe backbone distortion as shown by molecular modeling (Figure 2, top part). Furthermore, the intrastrand base stacking can be preserved if the first C·GxG triplet in reverse Hoogsteen configuration is flipped out from anti to syn glycosidic conformation (Figure 2, bottom part). It must be noted, however, that although the energetics of such a triplex (CG2 in Table 2) appear significantly favorable compared to the triplex where the guanine remains in anti glycosidic conformation, which could explain the experimental results, this base-flipping hypothesis has not been proven in the present work. Further structural studies are required to validate this model.

It is gratifying that the predictions of molecular modeling are, in general, qualitatively validated by the experimental data. This could be ascribed to the fact that the main problems involved in the computational optimization of a switch triplex are the geometrical and steric factors which are relatively easily solved by conformational energy minimization, despite the approximations made in the calculation and in the analysis (see Materials and Methods). Similar reasoning could explain why the detailed ranking of triplexes containing an additional nucleotide facing the 5′CpG3′ junction (see Table 1) was not fully validated by the experimental data (see Figure 4A).

The present work shows that the stability of an optimized switch triplex containing a 5′GpC3′ junction step is about 3−5 times higher than that containing a 5′CpG3′ junction step. This observation is consistent with previous findings that the switch triplexes appear to be more stable at the 5′RpY3′ junction step than at the 5′YpR3′ ones (*12*, *16*, *33*, *35*−*38*), a phenomenon which can be explained by the structure at the 5′RpY3′ junction where energetically favored base stacking could be possible at the strand crossover level within the switch third-strand oligonucleotide.

It has been found that the replacement of the antiparallel (G,T)-motif part of the switch oligonucleotide which recognizes box II through reverse Hoogsteen hydrogen bonding by the corresponding (G,A)-motif improves slightly the binding at room temperature and above. It should be noted that a destabilization was observed at 4 °C (data not shown). This was due to self-association of the (G,T,A)-containing oligomers.

It is noted that the current approach exploited a particular sequence context where an oligopurine strand contains quite lengthy A and G tracts with few 5′ApG3′ and 5′GpA3′ steps so that a (G,T)-containing third-strand oligonucleotide can form an unusual parallel (G,T)-motif triple helix. This sequence limitation can be overcome by combining together two parallel (C,T)- and (G,T)-motifs, provided the C·GxG base triplets were preserved at the junction.

As a concluding remark, it is instructive to consider more carefully the two types of junction steps in alternating oligopurine/oligopyrimidine sequences, i.e., 5′RpY3′ and 5′YpR3′. Each consists of three distinct subtypes. They are 5′ApT3′, 5′GpC3′, and 5′GpT3′ (which is equivalent to 5′ApC3′) for the 5′RpY3′ junction steps, and 5′TpA3′, 5′CpG3′, and 5′TpG3′ (which is also equivalent to 5′CpA3′) for the 5′YpR3′ junctions. Optimization of the third strand crossover at the 5′CpG3′ and 5′GpC3′ steps is described in the present work. Other experiments have also been carried out aimed at optimizing the third-strand crossover at different junctions. A comprehensive knowledge of how to cope with

all six junction steps (referred to as a "switch code") would theoretically open the door to extend the range of DNA recognition by switch triple helix formation.

ACKNOWLEDGMENT

We thank Dr. Richard Lavery for his advice on molecular modeling and for the use of the JUMNA program package. Dr. Eric Renoult and M. Bruno Faucon are also acknowledged for their preliminary contributions on this work.

REFERENCES

1. Le Doan, T., Perrouault, L., Praseuth, D., Habhoub, N., Decout, J.-L., Thuong, N. T., Lhomme, J., and Hélène, C. (1987) *Nucleic Acids Res. 15*, 7749−7760.
2. Moser, H. E., and Dervan, P. B. (1987) *Science 238*, 645−650.
3. Faucon, B., Mergny, J.-L., and Hélène, C. (1996) *Nucleic Acids Res. 24*, 3181−3188.
4. Noonberg, S. B., Francois, J. C., Praseuth, D., Guieysse-Peugeot, A. L., Lacoste, J., Garestier, T., and Hélène, C. (1995) *Nucleic Acids Res. 23*, 4042−4049.
5. Noonberg, S. B., Francois, J. C., Garestier, T., and Hélène, C. (1995) *Nucleic Acids Res. 23*, 1956−1963.
6. Svinarchuk, F., Debin, A., Bertrand, J. R., and Malvy, C. (1996) *Nucleic Acids Res. 24*, 295−302.
7. Svinarchuk, F., Bertrand, J. R., and Malvy, C. (1994) *Nucleic Acids Res. 22*, 3742−3747.
8. Svinarchuk, F., Paoletti, J., and Malvy, C. (1995) *J. Biol. Chem. 270*, 14068−14071.
9. Cheng, A. J., and Vandyke, M. W. (1994) *Nucleic Acids Res. 22*, 4742−4747.
10. Clarenc, J. P., Lebleu, B., and Leonetti, J. P. (1994) *Nucleosides Nucleotides 13*, 799−809.
11. Cooney, M., Czernuszewicz, G., Postel, E. H., Flint, S. J., and Hogan, M. E. (1988) *Science 241*, 456−459.
12. de Bizemont, T., Duval valentin, G., Sun, J. S., Bisagni, E., Garestier, T., and Hélène, C. (1996) *Nucleic Acids Res. 24*, 1136−1143.
13. Durland, R. H., Kessler, D. J., Gunnell, S., Duvic, M., Pettitt, B. M., and Hogan, M. E. (1991) *Biochemistry 30*, 9246−9255.
14. Olivas, W. M., and Maher, L. J. (1995) *Biochemistry 34*, 278−284.
15. Roy, C. (1993) *Nucleic Acids Res. 21*, 2845−2852.
16. Sun, J. S., de Bizemont, T., Duval-Valentin, G., Montenay-Garestier, T., and Hélène, C. (1991) *C. R. Acad. Sci., Ser. III 313*, 585−590.
17. Beal, P. A., and Dervan, P. B. (1992) *Nucleic Acids Res. 20*, 2773−2776.
18. Baumann, J., and Fayer, M. D. (1986) *J. Chem. Phys. 85*, 4087−4107.
19. Griffin, L. C., and Dervan, P. B. (1989) *Science 245*, 967−971.
20. Kiessling, L. L., and Dervan, P. B. (1992) *Biochemistry 31*, 2829−2834.
21. Mergny, J. L., Sun, J. S., Rougée, M., Montenay-Garestier, T., Barcelo, F., Chomilier, J., and Hélène, C. (1991) *Biochemistry 30*, 9791−9798.
22. Stilz, H. U., and Dervan, P. B. (1993) *Biochemistry 32*, 2177−2185.
23. Yoon, K., Hobbs, C. A., Koch, J., Sardaro, M., Kutny, R., and Weis, A. L. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 3840−3844.
24. Distefano, M. D., Shin, J. A., and Dervan, P. B. (1991) *J. Am. Chem. Soc. 113*, 5901−5902.
25. Distefano, M. D., and Dervan, P. B. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 1179−1183.
26. Zhou, B. W., Puga, E., Sun, J. S., Garestier, T., and Hélène, C. (1995) *J. Am. Chem. Soc. 117*, 10425−10428.
27. Froehler, B. C., Terhorst, T., Shaw, J. P., and McCurdy, S. N. (1992) *Biochemistry 31*, 1603−1609.
28. Horne, D. A., and Dervan, P. B. (1990) *J. Am. Chem. Soc. 112*, 2435−2437.
29. McCurdy, S., Moulds, C., and Froehler, B. (1991) *Nucleosides Nucleotides 10*, 287−290.
30. Ono, A., Chen, C. N., and Kan, L. S. (1991) *Biochemistry 30*, 9914−9921.
31. Ouali, M., Bouziane, M., Ketterle, C., Gabarroarpa, J., Auclair, C., and Lebret, M. (1996) *J. Biomol. Struct. Dyn. 13*, 977.
32. Zhou, B.-W., Marchand, C., Asseline, U., Thuong, N. T., Sun, J., Garestier, T., and Hélène, C. (1995) *Bioconjugate Chem. 6*, 516−523.
33. Beal, P. A., and Dervan, P. B. (1992) *J. Am. Chem. Soc. 114*, 4976−4982.
34. Bouziane, M., Cherny, D. I., Mouscadet, J. F., and Auclair, C. (1996) *J. Biol. Chem. 271*, 10359−10364.
35. Jayasena, S. D., and Johnston, B. H. (1992) *Biochemistry 31*, 320−327.
36. Jayasena, S. D., and Johnston, B. H. (1993) *Biochemistry 32*, 2800−2807.
37. Olivas, W. M., and Maher, L. J. (1996) *Nucleic Acids Res. 24*, 1758−1764.
38. Olivas, W. M., and Maher, L. J., III (1994) *Biochemistry 33*, 983−991.
39. Lavery, R., and Sklenar, H. (1988) *J. Biomol. Struct. Dyn. 6*, 63−91.
40. Lavery, R., Sklenar, H., Zakrzewska, K., and Pullman, B. (1986) *J. Biomol. Struct. Dynam. 3*, 929−1014.
41. Ouali, M., Letellier, R., Sun, J. S., Akhebat, A., Adnet, F., Liquier, J., and Taillandier, E. (1993) *J. Am. Chem. Soc. 115*, 4264−4270.
42. Ouali, M., Letellier, R., Adnet, F., Liquier, J., Sun, J. S., Lavery, R., and Taillandier, E. (1993) *Biochemistry 32*, 2098−2103.
43. Cantor, C. R., and Warshaw, M. M. (1970) *Biopolymers 9*, 1059−1077.
44. Marchand, C., Bailly, C., Hung Nguyen, C., Bisagni, E., Garestier, T., Helene, C., and Waring, M. J. (1996) *Biochemistry 35*, 5022−5032.
45. Sun, J.-S. (1995) in *Modelling of Biomolecular Structures and Mechanisms* (Pullman, A., et al., Eds.) pp 267−288, Kluwer Academic Publishers, Dordrechts, The Netherlands.

BI980618+