The Ff Gene 5 Single-Stranded DNA-Binding Protein Binds to the Transiently Folded Form of an Intramolecular G-Quadruplex†

Jin-Der Wen and Donald M. Gray*

Department of Molecular and Cell Biology, The University of Texas at Dallas, Box 830688, Richardson, Texas 75083-0688

Received April 12, 2002

ABSTRACT: The Ff gene 5 protein (g5p) is classified as a single-stranded DNA-binding protein. However, we previously showed that g5p binds with high affinity to a SELEX-selected G-rich 58-mer DNA oligomer, I-3, that forms an intramolecular G-quadruplex [Wen, J.-D., Gray, C. W., and Gray, D. M. (2001) *Biochemistry 40*, 9300−9310]. In 200 mM NaCl at 37 °C, g5p binds to I-3 in two stages, the first stage being the formation of a discrete intermediate complex that appears to be a precursor to a saturated g5p·I-3 complex. For the present paper, CD spectroscopy and DMS methylation techniques were used to investigate the binding of g5p to the I-3 oligomer and to the truncated 26-nucleotide core of the I-3 oligomer. The core sequence, called I-3c26, was d(GGGGTCAGGCTGGGGTTGTGCAGGTC). Results were the following: (1) The g5p binds in one stage to I-3c26 in 200 mM NaCl at 37 °C. (2) The intermediate complex of g5p·I-3 is formed by the binding of g5p to the core sequence. (3) G-quadruplex structures are maintained in both the g5p·I-3 and g5p·I-3c26 complexes, but the bound G-quadruplex structures are altered from their respective steady-state folded forms in 200 mM NaCl. (4) CD kinetics measurements showed that the I-3c26 quadruplex folds in two stages and that a transiently folded form is apparently the same as the altered structure to which g5p binds. (5) DMS methylation protection and interference experiments identified two guanines that are differentially involved in the steady-state folded and g5p-bound G-quadruplex structures. A model for a possible I-3c26 G-quadruplex structure is described.

The g5p[1] of the Ff phages binds to and saturates the nascent viral single-stranded genomic DNA to form a precursor for virion assembly (*1*). As a consequence of this biological function, g5p is known as a nonspecific, cooperative ssDNA binding protein. The g5p exists as a stable homodimer with a 2-fold rotational symmetry, and it binds to two antiparallel ssDNA strands (*2*−*5*). The number of nucleotides, *n* (the binding mode), bound per g5p monomer is dependent on the binding conditions. The $n = 4$ binding mode is dominant when g5p binds to genomic or purine-rich DNA at P/N ratios less than 0.25 (*6*, *7*).

The g5p can also bind at or adjacent to G-quadruplex structures (*8*, *9*). The g5p has been shown to repress translation of the Ff gene 2 mRNA by binding to a 16-mer leader sequence r($GU_5G_4CU_4C$) in the 5′ untranslated region of the mRNA (*10*). Kneale and colleagues have shown that this 16-mer leader sequence, or its DNA analogue, forms a four-stranded G-quadruplex structure through association of the central G_4 blocks. In the case of the 16-mer DNA d($GT_5G_4CT_4C$), the g5p preferentially binds to the structured form, as opposed to the single-stranded oligomer (*8*, *11*). Kneale and co-workers propose that tails of four strands, held together by G-quartets, are separated by the right distance to occupy the two symmetry-related binding sites on a g5p dimer and thus to favor g5p binding (*8*). In this model, the four DNA strands have to be antiparallel, which would be unusual for an intermolecular G-quadruplex.

Using SELEX to study specific sequences bound by the g5p, we identified a 58-mer ssDNA sequence (named I-3) that g5p binds with high affinity in 200 mM NaCl, pH 7.4, and 37 °C (*9*). The central, core 26 nucleotides of I-3 are G-rich, and CD measurements showed that I-3 indeed forms a G-quadruplex structure in 200 mM NaCl. Other measurements led to the conclusion that the I-3 G-quadruplex is an intramolecular structure. Nuclease S1 and end boundary experiments strongly suggested that g5p initially binds to the 26-mer core structure of I-3 and that the protein saturates the two flanking 16-mer tails only at a later stage of binding. The formation of a G-quadruplex site for initial g5p binding is probably responsible for the selection of the I-3 sequence as having high-binding affinity for g5p under the selection conditions.

Guanine-rich nucleic acid motifs that can form G-quadruplexes in vitro are found in naturally occurring sequences, such as telomeric sequences (*12*, *13*), IgG switch regions (*14*), fragile X syndrome nucleotide repeats (*15*), and

† This work was performed by J.-D.W. in partial fulfillment of the requirements for the Ph.D. degree in the Department of Molecular and Cell Biology, The University of Texas at Dallas. Support was provided by grants from the Robert A. Welch Foundation (AT-503) and the Texas Advanced Technology Program (009741-0021-1999).

* To whom correspondence should be addressed. Phone: (972) 883-2513. Fax: (972) 883-2409. E-mail: dongray@utdallas.edu.

[1] Abbreviations: CD, circular dichroism; DMS, dimethyl sulfate; EMSA, electrophoretic mobility shift assay; Ff phages, three closely related filamentous viruses f1, fd, and M13 that specifically infect F^+ strains of *Escherichia coli*; g5p, gene 5 protein; P/N, [protein monomer]/[nucleotide] molar ratio; SELEX, *s*ystematic *e*volution of *l*igands by *ex*ponential enrichment; ssDNA, single-stranded DNA; SVD, singular value decomposition; TAE buffer, 40 mM Tris-acetate, pH 8.3, and 1 mM EDTA; TBE buffer, 90 mM Tris-borate, pH 8.3, and 2 mM EDTA; TE buffer, 10 mM Tris-HCl, pH 7.4, and 1 mM EDTA; T_m, melting temperature.

10.1021/bi020276e CCC: $22.00
Published on Web 08/30/2002

the control region of the *c-myc* oncogene (*16*). The formation of G-quadruplex structures may play a positive role in some cellular functions but could hinder other functions such as replication and recombination (*17*). Although the presence of G-quadruplexes has not been demonstrated in vivo, their importance in biological systems is supported by the number of proteins that interact with them in vitro. These proteins include: (A) molecular chaperones that promote G-quadruplex formation (*18*−*21*), (B) nucleases that have G-quadruplex-dependent activities (*22*−*24*), (C) helicases that unwind G-quadruplexes more efficiently than they unwind double helices (*25*−*28*), and (D) other proteins (*29*−*31*) that have high affinity for binding to G-quadruplexes but may not have enzymatic activities.

CD spectroscopy has been widely used to diagnose the existence of G-quadruplexes. The structure of a quadruplex adopted by four parallel strands of d(TGGGGT) is a right-handed helix with a 4-fold symmetry and with all guanosines having an *anti* conformation of the glycosyl bonds (*32*). The CD spectrum of such a G-quadruplex exhibits a positive band at ∼265 nm and a small negative band at ∼245 nm (*11*, *33*). In contrast, the structure of an antiparallel quadruplex formed by two folded strands of d(GGGGTTTTGGGG) is also a right-handed helix, but the guanosines have alternating *syn* and *anti* glycosyl conformations along each strand (*34*−*36*). The CD spectrum of this sequence exhibits a positive band at ∼295 nm, a negative band at 260−265 nm, and a small positive band at ∼245 nm (*37*, *38*). When the glycosyl bonds of the guanosines of antiparallel quadruplexes have alternating *syn* and *anti* conformations along each strand, the G-quartet polarity (i.e., the direction of proton donors and acceptors within a given quartet) also alternates (*34*, *35*, *39*, *40*). Parallel quadruplexes have nonalternating G-quartet polarity (*32*). Therefore, given the sensitivity of CD measurements to base stacking geometry, it is not surprising that parallel and antiparallel G-quadruplexes have different CD spectra (*38*, *41*). In general, a positive band at ∼295 nm is associated with an antiparallel strand arrangement, although such G-quadruplexes can also have a positive band at 255−265 nm (*37*, *38*, *41*−*43*), possibly because the polarity of adjacent G-quartets in antiparallel G-quadruplexes does not always alternate (*39*, *40*). In addition, the CD spectra of antiparallel folds, and in particular the 295 nm CD band, may be influenced by the connecting loops (*44*, *45*). The most simple antiparallel G-quadruplex folds are of two basic types. Folds having only edgewise loops connecting adjacent strands are designated as "chair" forms, while those having a diagonal loop are designated as "basket" forms (*17*).

In the present paper, we use CD spectroscopy to analyze the G-quadruplex structures formed by I-3 and a truncated I-3c26 sequence. (I-3c26 consists of the central 26 nucleotides of I-3.) CD was also used to characterize the kinetics of I-3c26 folding induced by NaCl and by g5p binding. The particular guanines that are involved in the free I-3c26 and g5p-bound I-3 G-quadruplex structures were determined by DMS methylation experiments. The G-quartets appear to include bases other than G. The present work establishes that g5p indeed binds to, and perturbs, an intramolecular G-quadruplex fold. Moreover, the perturbation by g5p binding reestablishes a transiently folded form of the quadruplex.

EXPERIMENTAL PROCEDURES

DNA Sequences. The DNA sequences I-3, I-3c26, I-3c27, and G-8c26 used in this study were purchased from Oligos Etc. (Wilsonville, OR). Sequences were named for the clones from which they were derived. The 58-mer I-3 sequence was 5′-d(CGGGATCCAACGTTTT-**GGGG**TCA**GG**CT**GGGG**T-T**GTG**CAGGTC-AAGAGGCAGAATTCGC)-3′; the I-3c26 sequence was the underlined central sequence of 26 nucleotides; the 27-mer I-3c27 sequence was identical to I-3c26 with the addition of one nucleotide "T" at the 5′ end. The 26-mer G-8c26 sequence was 5′-d(TA**GGG**CA**GGGG**-TCGTC**GGG**TTA**GGG**C)-3′. Blocks of G's (plus an important GTG sequence in I-3) are shown in bold type in the above sequences. The two 16-mer tails of the I-3 sequence were the primer binding sites originally needed for the SELEX procedure (*9*).

g5p. The g5p was isolated from *Escherichia coli* transformed with an Ff gene 5-containing plasmid and purified using DNA cellulose affinity plus Sephadex G75 sizing columns as in previous work (*7*, *46*). The purity was >99% as determined by quantitation of Coomassie Blue stained bands on sodium dodecyl sulfate-polyacrylamide gels.

EMSA. EMSA for g5p titrations was performed as in previous work (*9*). In brief, 1 μM I-3 or 2.2 μM I-3c26 was ^{32}P-labeled and incubated with 0−21 μM g5p in TE buffer containing 200 mM NaCl at 37 °C for 15−30 min. Mixtures were subjected to electrophoresis on 2.5% (w/v) low-melting agarose gels (BioWhittaker Molecular Applications, Rockland, ME) in TAE buffer (40 mM Tris-acetate, pH 8.3, 1 mM EDTA). The gels were fixed, dried, and exposed to a storage phosphor screen (Molecular Dynamics, Sunnyvale, CA).

CD and UV Measurements. CD spectra were measured using a Jasco Model J715 spectropolarimeter (Jasco, Inc., Easton, MD) as in previous work (*46*). The CD spectra were smoothed by the method of fast Fourier transformation (Jasco) and plotted at 1 nm intervals as molar CD ($\epsilon_L - \epsilon_R$) in units of M^{-1} cm^{-1}, per mole of nucleotide.

For kinetics experiments, the CD was monitored at 255 or 292 nm with a bandwidth of 5 nm. Data were collected at time intervals of 0.5 or 1 s before and after manual addition of NaCl or g5p to the sample in a cuvette. Time-dependent molar CD ($\epsilon_L - \epsilon_R$) changes were fitted using SigmaPlot 5.0 (SPSS Inc., Chicago, IL) with a single-exponential function $f(t) = C + A \exp(-kt)$ or a double-exponential function $f(t) = C + A_1 \exp(-k_1 t) + A_2 \exp(-k_2 t)$, in which C was a constant, A, A_1, and A_2 were the amplitudes, and k, k_1, and k_2 were the rate constants.

UV melting profiles of oligonucleotides were measured in an Olis-modified Cary model 14 spectrophotometer (OLIS, Inc., Bogart, GA). Samples were usually melted from 5 to 90 °C at 1 °C temperature intervals, with a 3 min incubation time at each temperature interval before the acquisition of absorbance data at 260 or 255 nm.

SVD. Singular value decomposition was performed as described by Hashem et al. (*47*).

DMS Methylation. I-3c26 and I-3c27 were ^{32}P-labeled at the 5′ or 3′ ends, and I-3 was labeled at the 5′ end, as described previously (*9*).

(A) Methylation Protection Experiments. Radioactive labeled oligonucleotide (250−600 pmol) was allowed to fold

into G-quadruplexes in TE buffer (10 mM Tris-HCl, pH 7.4, 1 mM EDTA) containing 200 mM NaCl by heating at 37 °C for >30 min. DMS was added to give a final concentration of 0.2% (v/v) in a total volume of 200 μL. The methylation reaction was performed at room temperature for 10 min and stopped by addition of 50 μL of 1 M β-mercaptoethanol and 1.5 M sodium acetate (pH 7.0). Oligonucleotides were recovered by ethanol precipitation prior to piperidine treatment.

(B) Methylation Interference Experiments. Randomly methylated oligonucleotides were obtained by the same DMS treatment as for the methylation protection experiments but in the absence of NaCl. (i) To determine the sequence positions that interfered with the formation of G-quadruplex structures in I-3c26 or I-3c27, the methylated oligonucleotide samples were resuspended in TE buffer containing 200 mM NaCl, folded by heating at 37 °C for >30 min, and subjected to electrophoresis at 4 °C in 15% (w/v) native polyacrylamide gels in TBE buffer (90 mM Tris-borate, pH 8.3, 2 mM EDTA) containing 100 mM NaCl. Gel slices corresponding to a higher mobility band (structured forms) and a lower mobility band (unstructured forms) were isolated. DNAs were eluted from these bands and ethanol-precipitated prior to piperidine treatment. (ii) To determine the sequence positions that interfered with the formation of a g5p•I-3 initiation complex, ~9 μM methylated I-3 was incubated with 22 μM g5p in TE buffer plus 200 mM NaCl at 37 °C for 15 min. The mixture was subjected to electrophoresis on a 2.5% agarose gel as described above for EMSA experiments, and a gel slice with the initiation complex was isolated. The DNA from the initiation complex was eluted from the gel slice and ethanol-precipitated prior to piperidine treatment.

Piperidine treatment was performed as per Maxam and Gilbert (*48*) except that the cleavage reaction at 90 °C was reduced to 15 min to minimize background cleavage. Piperidine treatment causes strand cleavage at guanosines that are methylated at the guanine N^7 position. Sequence fragments were separated on 12% or 20% (w/v) denaturing polyacrylamide gels in TBE buffer. Gels were dried and exposed to a storage phosphor screen for visualization.

RESULTS

EMSA of Titrations of I-3 and I-3c26 with g5p. As shown in Figure 1A, titration of I-3 with g5p in 200 mM NaCl involved two stages of binding. An intermediate, unsaturated complex (initiation complex) appeared at low P/N ratios, followed by the appearance of a saturated complex at P/N ratios above about 0.06. In contrast, as seen in Figure 1B, titration of I-3c26 showed one stage of binding, and only one discrete band of complex was formed. These results support the idea of sequential binding of g5p, first to the central segment and then to the two 16-mer tails of I-3 to form the initiation and saturated complexes, respectively, as previously proposed (*9*).

A spread of lower mobility complexes above the discrete band of the g5p•I-3c26 complex occurred at P/N ratios above 0.3 (most evident in Figure 1B, lane 10). This less distinct distribution of complexes was observed only for oversaturating amounts of g5p and thus was not equivalent to the second stage of g5p binding to I-3. This distribution was gradually

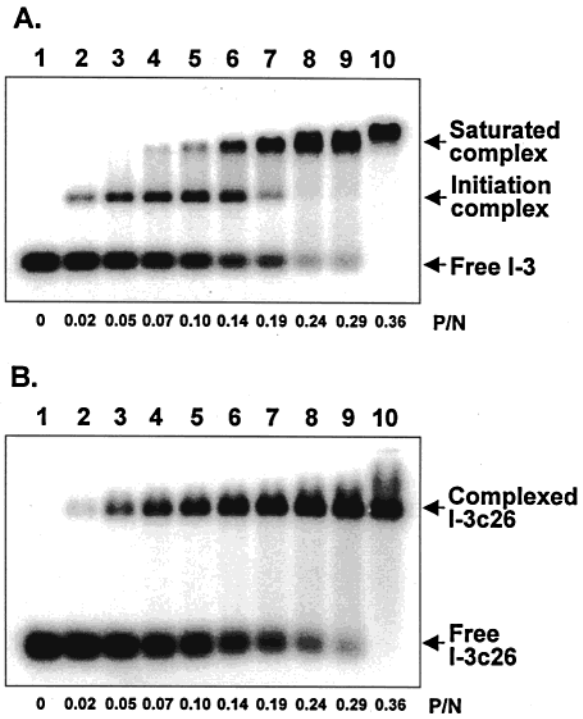

FIGURE 1: EMSA of I-3 and I-3c26 titrated with g5p. (A) EMSA of I-3 titrated with g5p. ^{32}P-Labeled I-3 (1 μM in strand or 58 μM in nucleotide) was titrated with increasing concentrations of g5p at 37 °C in TE buffer with 200 mM NaCl. Samples were electrophoresed on a 2.5% agarose gel in TAE buffer. The gel was fixed, dried, and analyzed as described in Experimental Procedures. P/N ([protein monomer]/[nucleotide]) molar ratios are shown at the bottom of the panel. The protein concentration at a P/N ratio of 0.25 was theoretically sufficient to saturate the oligonucleotide in the $n = 4$ binding mode. (B) EMSA of I-3c26 titrated with g5p. ^{32}P-Labeled I-3c26 (2.23 μM in strand or 58 μM in nucleotide) was titrated with g5p and electrophoresed as in panel A.

converted to a discrete band during a longer incubation (data not shown). It was not observed for complexes formed with I-3 (Figure 1A). In addition, CD kinetic evidence suggested that there was a structural rearrangement of I-3c26 in the complex with g5p, as will be described below. For these reasons, we believe that the spread of lower mobility complexes in the right-most lanes of Figure 1B likely reflects the state of complexes in the sample as opposed to being the result of a reequilibration of the structures during exposure to the TAE running buffer.

Strand Stoichiometry of I-3c26. In general, G-rich sequences may form intramolecular and/or intermolecular G-quadruplexes that can be distinguished by gel electrophoresis (*49*). However, I-3 appears to exist only as an intramolecular structure in 200 mM NaCl (*9*). In additional experiments, we found that I-3c26 also exists as an intramolecular structure at the same ionic strength. Samples of I-3c26 prepared in 200 mM NaCl were heated at 95 °C for 5 min and slowly cooled to room temperature. Samples all migrated as single bands on a 15% native polyacrylamide gel containing 100 mM NaCl regardless of DNA concentration (0.1−10 μM) or the incubation time (up to 18 h) after cooling (data not shown). The intramolecular nature of the I-3c26 G-quadruplex was further confirmed by CD kinetics experiments (see below).

CD of I-3 and I-3c26 Titrated with NaCl. Figure 2A shows CD spectra obtained during a titration of I-3 with NaCl at

Table 1: Melting Temperatures (T_m) and Percent Hyperchromicities (% H) of the Dominant Transitions in the Melting Profiles of I-3, I-3c26, and G-8c26 at 10 and 200 mM NaCl

	I-3 at [NaCl]		I-3c26 at [NaCl]		G-8c26 at [NaCl]	
	10 mM	200 mM	10 mM	200 mM	10 mM	200 mM
T_m (°C)[a]	19.1 ± 1.3	37.7 ± 0.5[b]	30.9 ± 0.2	54.8 ± 0.4	34.4 ± 0.3	60.5 ± 0.5
% H at T_m[c]	7.0 ± 0.5	9.9 ± 0.6	6.1 ± 0.1	9.3 ± 0.5	~3.7	~6.1

[a] T_m values were the mean ± standard deviation from derivatives of at least three melting profiles. Absorbance measurements of I-3 and I-3c26 were made at 260 nm, and measurements of G-8c26 were made at 255 nm. The buffer was 10 mM Tris-HCl, pH 7.4. [b] There was an additional shoulder above 50 °C in melting profiles of I-3 at 200 mM NaCl. [c] Percent hyperchromicities were determined as the percent increase in optical density at 260 or 255 nm when the temperature was at the T_m, relative to the lowest temperature (~5 °C).

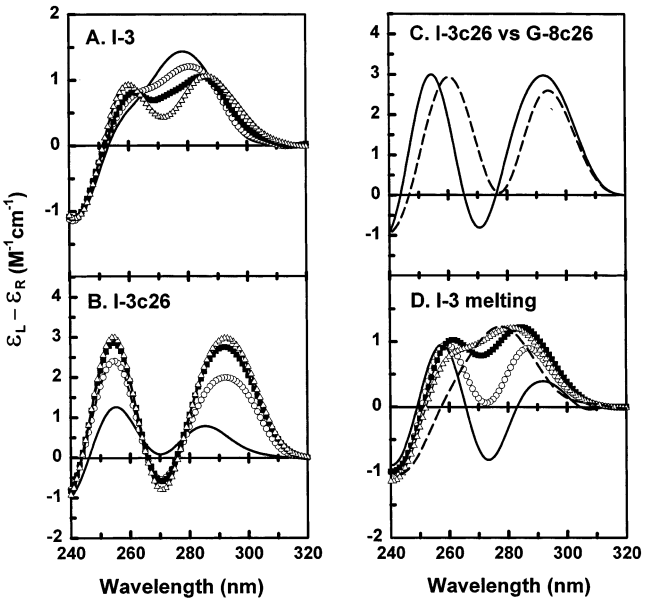

FIGURE 2: CD spectra of I-3 and I-3c26 titrated with NaCl, comparison of I-3c26 and G-8c26 spectra, and temperature-dependent spectra of I-3. (A) Representative CD spectra taken during the titration of I-3 with NaCl at 37 °C. CD spectra are shown at NaCl concentrations of 10 mM (—), 50 mM (○), 100 mM (■), and 200 mM (△). (B) Representative CD spectra taken during the titration of I-3c26 with NaCl at 37 °C. CD spectra are shown at the same NaCl concentrations as in panel A. (C) CD spectra of I-3c26 (—) and G-8c26 (- - -). Both were taken in 200 mM NaCl at 37 °C. (D) CD spectra of I-3 in 200 mM NaCl as the temperature was increased from 5 to 90 °C. CD spectra are shown at 5 °C (—), 30 °C (○), 40 °C (■), 50 °C (△), and 90 °C (- - -). CD spectra for this figure and Figures 3 and 4 were all taken with samples in 10 mM Tris-HCl, pH 7.4, plus the specified concentrations of NaCl and at a temperature of 37 °C, unless otherwise specified.

37 °C. At 10 mM NaCl, I-3 showed a positive CD band at 278 nm. This CD spectrum represented an unstructured, single-stranded form of I-3, since it was essentially identical to the spectrum of I-3 at 90 °C (Figure 2D). As the NaCl concentration was gradually increased to 200 mM, the 278 nm band of I-3 was replaced by two bands at 260 and 286 nm, indicative of the formation of a G-quadruplex (*37*, *38*, *42*, *43*). Thus, a structural change of I-3 was induced by an increase in NaCl concentration at 37 °C.

In the case of I-3c26, two positive CD bands at 255 and 285−292 nm, analogous to those in the spectrum of I-3 at high salt concentration, were already present at 10 mM NaCl and 37 °C (Figure 2B). These positive bands were further enhanced as the salt concentration increased. The presence of two similar positive CD bands at long wavelengths suggested that I-3c26 and I-3 in 200 mM NaCl shared some common structural features.

We further compared the CD spectrum of I-3c26 with that of G-8c26. G-8c26 consisted of the central 26 nucleotides of another G-rich sequence (G-8) that was also SELEX-selected by g5p binding (*9*). The I-3c26 and G-8c26 sequences were alike in having blocks of two-to-four G's, a common feature for most of the SELEX-selected sequences (*9*). In the case of G-8c26, there were at least three G's in each of four blocks, making it likely that a G-quadruplex would consist of three G-quartets. This oligomer showed two positive CD bands at 260 and 294 nm, similar to those of I-3c26 in wavelengths and magnitudes, under the same conditions of 200 mM NaCl and 37 °C (Figure 2C). Therefore, G-8c26 and I-3/I-3c26 probably formed related G-quadruplex structures.

Melting Temperatures of I-3, I-3c26, and G-8c26. Melting temperatures of I-3, I-3c26, and G-8c26 at low (10 mM) and high (200 mM) NaCl concentrations were obtained from the derivatives of UV absorption melting profiles at 260 or 255 nm. The T_m values and the hyperchromicities at the T_m are summarized in Table 1. In the cases of I-3c26 and G-8c26, only one obvious cooperative transition was observed. In 200 mM NaCl, the major transition in the melting profile of I-3 (37.7 °C) was at a lower temperature than that of I-3c26 (54.8 °C). Moreover, the CD spectral changes during the melting of I-3 in 200 mM NaCl indicated that the major transition in the absorption profile was not primarily due to the melting of its G-quadruplex core. As shown in Figure 2D, the CD band at long wavelengths increased in magnitude until 40 °C and then, together with the band at 260 nm, melted with a T_m above 50 °C. In 200 mM NaCl, a shoulder was also evident above 50 °C in the melting profile of I-3 that coincided with the loss of the 260 nm CD band of the quadruplex. Therefore, it appeared that the core structure of I-3 melted with a T_m close to that of the isolated I-3c26 core sequence in 200 mM NaCl. We reason that the apparent T_m of 37.7 °C for I-3 was mainly due to the disassociation of extraneous base pairing involving the 16-mer tails.

The T_m values of the 26-mer sequences were increased by 24−26 °C when the NaCl concentration was raised from 10 to 200 mM, consistent with an enhancement of CD bands when salt was added at 37 °C (see Figure 2B). The T_m values for G-8c26 were 4−6 °C higher than those for I-3c26, as might be expected since the former sequence can form a more homogeneous array of G-quartets.

CD of I-3 and I-3c26 Titrated with g5p. Figure 3A shows CD spectra obtained during a titration of I-3 with g5p at 37 °C in the presence of 200 mM NaCl. The CD spectrum of I-3 changed in two stages during the titration. First, the CD band at 260 nm increased with increasing P/N ratios up to

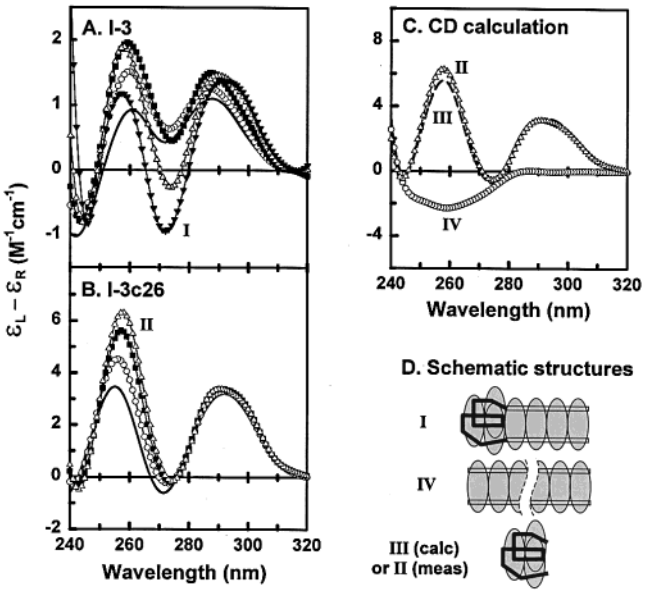

FIGURE 3: CD spectra of I-3 and I-3c26 titrated with g5p and comparison of spectra calculated for the g5p-bound core of I-3 and measured for g5p-saturated I-3c26. (A) Representative CD spectra taken during the titration of I-3 with g5p in 200 mM NaCl at 37 °C. CD spectra are shown at P/N ratios of 0 (—), 0.02 (○), 0.06 (■), 0.14 (△), and 0.29 (▼). Spectrum I is at the highest P/N ratio. (B) Representative CD spectra taken during the titration of I-3c26 with g5p in 200 mM NaCl at 37 °C. CD spectra are shown at P/N ratios of 0 (—), 0.05 (○), 0.11 (■), and 0.15 (△). Spectrum II is at the highest P/N ratio. (C) The measured CD spectrum for the g5p·I-3c26 complex (△, spectrum II from panel B), the calculated CD spectrum for the g5p-saturated core of I-3 (---, spectrum III), and a reference measured spectrum for the g5p-saturated single strands of I-3 in 10 mM NaCl (○, spectrum IV). Spectrum III was calculated as follows: CD(spectrum III) = [58 CD(spectrum I) − 32 CD(spectrum IV)]/26. See text. (D) Schematic structures of g5p·I-3 and g5p·I-3c26 complexes corresponding to the CD spectra labeled as I, IV, and III (calc) or II (meas). Solid lines represent structured DNA sequences, open lines represent unstructured ones, and ovals represent g5p dimers.

P/N = 0.06. Second, the CD at wavelengths centered at about 270 nm decreased until the P/N ratio reached ~0.25. The band at 286−290 nm changed less dramatically during the additions of g5p. The two stages of CD change at 260 nm were consistent with the gel titration data in which two bands of complexes sequentially appeared as the P/N ratio increased (Figure 1A). The structural changes of I-3 during the first stage of the titration presumably involved its G-quadruplex core, since other interactions were essentially melted out at 37 °C (Figure 2D).

During the titration of I-3c26 with g5p, only one transition was detected in which the CD band at 255 nm increased (Figure 3B). The CD data during the titration of I-3c26 were also consistent with the gel titration data in that only one band of complex was formed (Figure 1B). The simplest interpretation of both titrations is that the first stage of CD changes during the titration of I-3 with g5p (Figure 3A) corresponded to the direct binding of protein to the central segment of 26 nucleotides. This interpretation was in agreement with footprinting and other published evidence that the two tails of I-3 were not involved in forming an unsaturated initiation complex with g5p (*9*). Furthermore, the fact that the two positive CD bands were maintained in the saturated g5p·I-3 complex (Figure 3A) suggested that the I-3 core structure was preserved during g5p binding.

The sensitivity of CD measurements to structural changes allowed us to further test whether the core structure of I-3 when complexed with g5p (to form a saturated complex) was similar to the structure of the separate g5p·I-3c26 complex. The structural similarity of these two complexes was tested as follows: A CD spectrum of the core structure of the I-3 complex with g5p was calculated by subtracting 32 parts of the CD (per mole of nucleotide) of g5p-saturated single-stranded I-3 in 10 mM NaCl (spectrum "IV" in Figure 3C) from 58 parts of the CD (also per mole of nucleotide) of the saturated I-3 in 200 mM NaCl (spectrum "I" in Figure 3A). This CD difference, divided by 26 to reduce the CD back to a molar nucleotide value, gave a calculated CD spectrum for the g5p-bound core structure of I-3 (spectrum "III" in Figure 3C). This calculation was based on the simple assumption that each nucleotide made the same contribution to the spectrum of the g5p-saturated single-stranded I-3 sequence (spectrum "IV").

As seen in Figure 3C, the result of this CD calculation (spectrum "III") was a spectrum that was remarkably close to the measured spectrum of the g5p·I-3c26 complex (spectrum "II"). Given the closeness of this calculation, plus the fact that major increases occurred in the 255−260 nm CD bands of both I-3 and I-3c26 during initial g5p binding, there must be a close similarity of the structures of the g5p·I-3c26 complex and the core of the saturated g5p·I-3. Both appear to contain the same arrangement of G-quadruplexes and to be similarly perturbed upon the initial binding of g5p.

Simulation of CD Changes Corresponding to EMSA Analysis of Titrated I-3. The above CD spectral comparison and calculation suggested that the core structure of I-3 was similar to the structure of I-3c26 in 200 mM NaCl, especially when complexed with g5p, and, therefore, that the 16-mer tails and the core of I-3 could be distinguished by CD spectroscopy. Since the initiation complex always coexisted with free I-3 and/or a saturated complex (see Figure 1A), its CD spectrum could not be directly measured. However, with the assumption that the CD effects of g5p binding to the core and tails of I-3 were additive, it was possible to calculate the CD spectrum of the initiation complex of g5p·I-3 formed in 200 mM NaCl. As shown in Figure 4A, a CD spectrum of the initiation complex (spectrum "V") was calculated by adding 32 parts of the molar CD spectrum of the single-stranded form of I-3 (spectrum "VI" in Figure 4A), as the spectral equivalent of two tails, to 26 parts of the molar CD spectrum of g5p·I-3c26 (spectrum "II" in Figures 3B and 4A) and then dividing by 58 to bring the total back to the scale of CD per mole of nucleotide. The result of this calculation enabled us to simulate the CD spectra that corresponded to the EMSA results of titrating I-3 with g5p (Figure 1A). That is, each simulated CD spectrum at a given P/N ratio was reconstructed from the spectra of the free I-3 (— in Figure 3A), the saturated complex (spectrum "I" in Figure 3A), and the calculated initiation complex (spectrum "V" in Figure 4A), weighted by their percentages from EMSA data, which are shown in Figure 4B. The results are plotted in Figure 4C, which indeed closely resembled the actual titration data in Figure 3A. To quantitatively compare the simulated and measured CD titration data, a titration curve of CD values at 260 nm was plotted as a function of P/N ratio for both the simulated and measured titrations. As

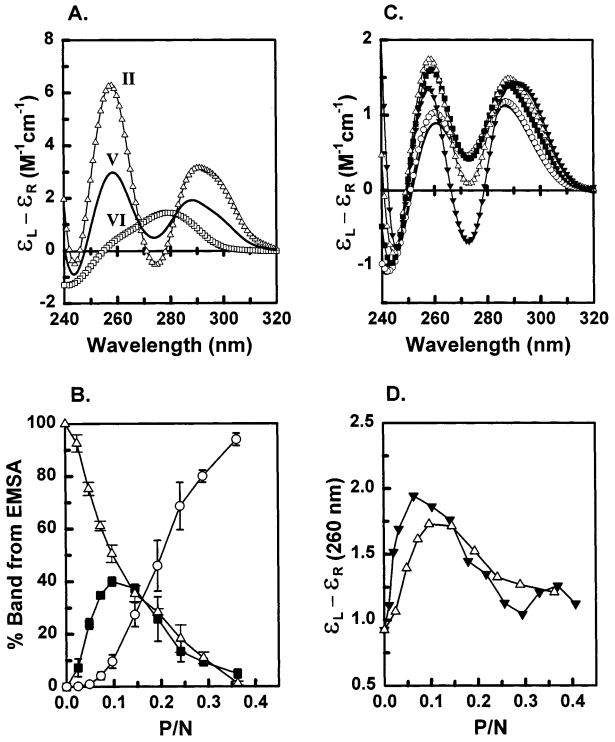

FIGURE 4: Simulation of CD spectra corresponding to EMSA results. (A) The measured CD spectrum for the g5p·I-3c26 complex (△, spectrum II from Figure 3B), the calculated CD spectrum for the g5p·I-3 initiation complex (—, spectrum V), and a reference measured spectrum for the free I-3 oligomer in 10 mM NaCl (□, spectrum VI). Spectrum V was calculated as follows: CD(spectrum V) = [26 CD(spectrum II) + 32 CD(spectrum VI)]/58. See text. (B) Quantitation of bands from EMSA analyses. Percentages of the saturated complex (○), initiation complex (■), and free DNA (△) at each P/N ratio are shown for titrations of I-3 with g5p in 200 mM NaCl at 37 °C. Data were averages from three EMSA experiments (e.g., as in Figure 1A), and error bars are standard deviations. (C) Simulation of CD spectra during a titration of I-3 with g5p. Representative CD spectra at P/N ratios of 0 (—), 0.02 (○), 0.07 (■), 0.14 (△), and 0.29 (▼) are shown. The CD spectrum at each P/N ratio was reconstructed from three component spectra: the measured spectrum for the saturated complex formed in 200 mM NaCl (i.e., spectrum I in Figure 3A), the measured spectrum for the free I-3 in 200 mM NaCl (i.e., spectrum VI), and the calculated spectrum for the initiation complex (i.e., spectrum V). The component spectra were weighted and summed according to the percentages shown in panel B. (D) Comparison of simulated (△) and measured (▼) CD titration curves as a function of P/N ratio.

shown in Figure 4D, the simulated titration curve (open triangles) lagged slightly behind the measured curve (closed triangles), but overall the two were close in shape. The difference could be due to approximations in the CD calculation of the initiation complex or could reflect an actual difference between the gel and solution (CD) titrations. Regardless of the source of this difference, the closeness of the simulated and measured CD titration curves provided additional evidence that the intermediate complexes resolved by gel electrophoresis and CD titrations were the same and were both formed via g5p binding to the core structure of I-3.

Singular Value Decomposition (SVD). SVD was applied to decompose the CD spectra obtained during the titration of I-3 with g5p in 200 mM NaCl (Figure 3A). SVD yielded only two significant basis vectors that are shown in Figure 5A, and the weights of the two major basis vectors as a function of P/N ratio are shown in Figure 5B. It was clear that basis vector 1 was the dominant contributor to CD changes at 260 and 286−290 nm (Figure 5A, open circles) and basis vector 2 was the contributor to the CD change at 270 nm (Figure 5A, closed squares) as the P/N ratio increased. In addition, basis vector 2 dramatically changed in magnitude and in sign during the titration (Figure 5B, closed squares), while basis vector 1 was relatively persistent (Figure 5B, open circles). Therefore, SVD unambiguously showed that the CD bands at 260 nm and 286−290 nm were maintained during the titration, supporting the CD analyses above that the G-quadruplex core structure of I-3 was maintained during g5p binding. Furthermore, the change in basis vector 2 was consistent with the binding of g5p to the two tails of I-3 as the P/N ratio increased above ~0.06.

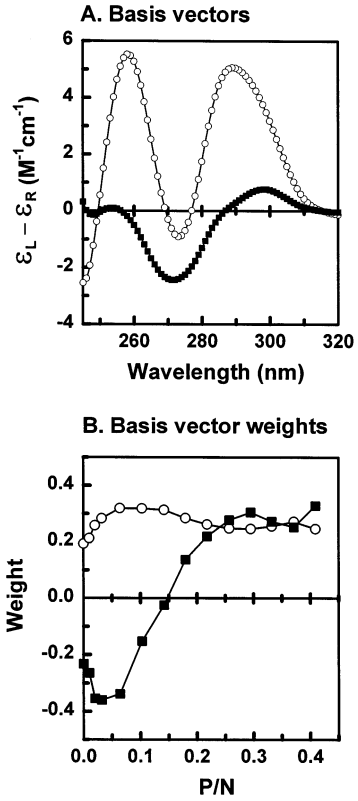

FIGURE 5: SVD of CD titration of I-3 with g5p. SVD was applied to decompose experimental CD spectra obtained during the titration of I-3 with g5p in 200 mM NaCl at 37 °C (see Figure 3A for representative spectra from such a titration). (A) Basis vectors. Only two significant CD basis vectors, 1 (○) and 2 (■), were obtained from SVD. (B) Basis vector weights. The weights for basis vectors 1 (○) and 2 (■) are plotted as a function of P/N ratio.

Two stages of g5p binding to I-3 were also evident from the SVD results. As shown in Figure 5B, the first stage occurred with P/N ratios up to ~0.06; in this stage the weight of basis vector 1 increased and that of basis vector 2 decreased. The second stage followed until the P/N ratio reached ~0.25; in the second stage, the weight of basis vector 1 decreased slightly and that of basis vector 2 increased dramatically. At P/N ratios >0.25, both vectors were basically unchanged, suggesting that no more g5p binding occurred.

Kinetics of I-3c26 Folding and g5p Binding. The kinetics of folding of I-3c26 could be monitored by CD spectroscopy, exploiting the significant increase in the CD bands of I-3c26 at 255 and 285−292 nm that accompanied the formation of

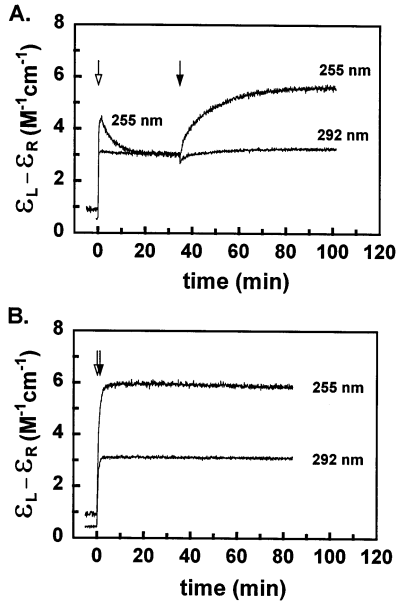

FIGURE 6: Kinetics of I-3c26 folding and g5p binding. (A) Time-dependent CD changes of I-3c26 after separate additions of 200 mM NaCl and g5p. I-3c26 (2.3−2.5 μM) was preequilibrated in 10 mM Tris-HCl buffer, pH 7.4, at 37 °C. An aliquot of 5 M NaCl was added to give a final concentration of 200 mM, and CD bands at 255 and 292 nm were monitored as a function of time at 1 s intervals. When the CD values had stopped changing, an aliquot of g5p was added to give a P/N ratio of 0.18, and monitoring of CD changes was resumed. The open and closed arrows indicate the respective times when NaCl and g5p were added. (B) Time-dependent CD changes of I-3c26 upon simultaneous addition of 200 mM NaCl and g5p. The same amounts of NaCl and g5p were added as in (A), except that g5p was added within 15 s of the addition of NaCl.

structured I-3c26 upon increasing the NaCl concentration (see Figure 2B). Data in Figure 6A show that the 292 nm CD band increased to its steady-state value within 1 min after a sudden increase in NaCl concentration to a final concentration of 200 mM, whereas the 255 nm band transiently increased to a peak CD value of about 4.5 M^{-1} cm^{-1} at 50−60 s and then slowly relaxed (Figure 6A, between two arrows). The time needed to reach the peak value at 255 nm, 50−60 s, was independent of I-3c26 concentration over the range of 1.1−3.9 μM, which indicated that the initial changes were relatively rapid first-order processes. In addition, the relaxation of the 255 nm band (60 s after the NaCl addition) was a first-order process as shown by the following: (a) The relaxation process was fitted by a single-exponential decay function with a rate constant $k = 0.20 \pm 0.02$ min^{-1}, and (b) the initial velocity of the relaxation was linearly proportional to the concentration of I-3c26 (data not shown). Therefore, the NaCl-induced CD change or structural change of I-3c26 occurred in at least two stages, both of which involved intramolecular rearrangements, supporting the results of gel electrophoresis experiments that this G-quadruplex was formed by intramolecular folding. Moreover, the structural features of I-3c26 corresponding to the 292 nm CD band formed immediately, and only those corresponding to 255 nm CD band underwent a further rearrangement.

CD kinetics was also used to study the I-3c26 structural rearrangement upon g5p binding. When g5p was added to I-3c26 that had reached a folded steady state at 200 mM NaCl and 37 °C (solid arrow in Figure 6A), the 255 nm CD band exponentially increased to a steady state, whereas the 292 nm band only underwent a minute fluctuation at the time of g5p addition. The rearrangement was basically first order, dominated by one rate constant. [The trace of the 255 nm band change was fitted by a double-exponential function $f(t) = C + A_1 \exp(-k_1 t) + A_2 \exp(-k_2 t)$. Experiments in which the DNA and protein concentrations were systematically varied indicated that the two rate constants k_1 and k_2 varied inversely with the P/N ratio; as the P/N ratio varied from 0.06 to 0.43, k_1 varied from 3.42 to 0.15 min^{-1} and k_2 varied from 0.17 to 0.02 min^{-1}.] As was the case of the slower of the two structural rearrangements of this oligomer that occurred with NaCl addition (Figure 6A, open arrow), only the CD at 255 nm was affected as the structure of I-3c26 was rearranged upon addition of g5p (Figure 6A, solid arrow).

A question was whether the increased 255 nm CD band upon addition of g5p to equilibrated I-3c26 meant that the I-3c26 structure returned to the same transiently folded form that occurred immediately after the addition of NaCl (Figure 6A, open arrow). To help to answer this question, g5p was added immediately after NaCl was added to the I-3c26 sample. As demonstrated in Figure 6B, the CD bands at 255 and 292 nm increased immediately and were maintained at their maximal values when g5p was added immediately after the NaCl. The steady-state values of the CD bands under these conditions were comparable in magnitude to those reached after g5p was added to the folded I-3c26 (compare magnitudes in panels A and B of Figure 6). Thus, it seemed likely that the transiently folded I-3c26 was in a conformation optimal for g5p binding and that the G-quadruplex structures of I-3c26 in its transiently folded form and in its g5p-complexed form were similar.

The actual structural differences that resulted in the separation of CD changes at 255 and 292 nm during I-3c26 folding and g5p binding are not known, but DMS methylation experiments (below) suggested that the differences involve specific guanines (G20 and/or G28).

DMS Methylation. To further confirm the existence of G-quadruplex structures in I-3/I-3c26 and to identify the guanines involved in the structures, DMS methylation protection and interference experiments were performed for 3′-end ^{32}P-labeled I-3c26 and I-3c27, and for 5′-end ^{32}P-labeled I-3. As seen in lane 3 of Figure 7A, 10 specific guanines (marked by arrowheads) were protected or partially protected from N^7 methylation in the folded form of I-3c27. Eight of these guanines were in three G blocks (G18-G20, G24-G25, and G29-G31) and two were the G's of a G34-T35-G36 sequence. These data leave little doubt that the G's were hydrogen-bonded in G-quartets. Moreover, methylation of the same guanines, except for G20, interfered with the subsequent formation of the I-3c27 structure upon addition of NaCl, as shown by the results of the methylation interference experiment in lanes 4 and 5 of Figure 7A (solid arrowheads). G20 was partially protected in the I-3c27 structure, but the methylation of the base did not seem to interfere with the formation of a G-quadruplex structure (Figure 7A, open arrowhead). The I-3c26 and I-3c27 sequences showed similar methylation patterns regardless of the end that was labeled (not shown). Only the data from I-3c27 labeled at the 3′ end are shown in Figure 7A because

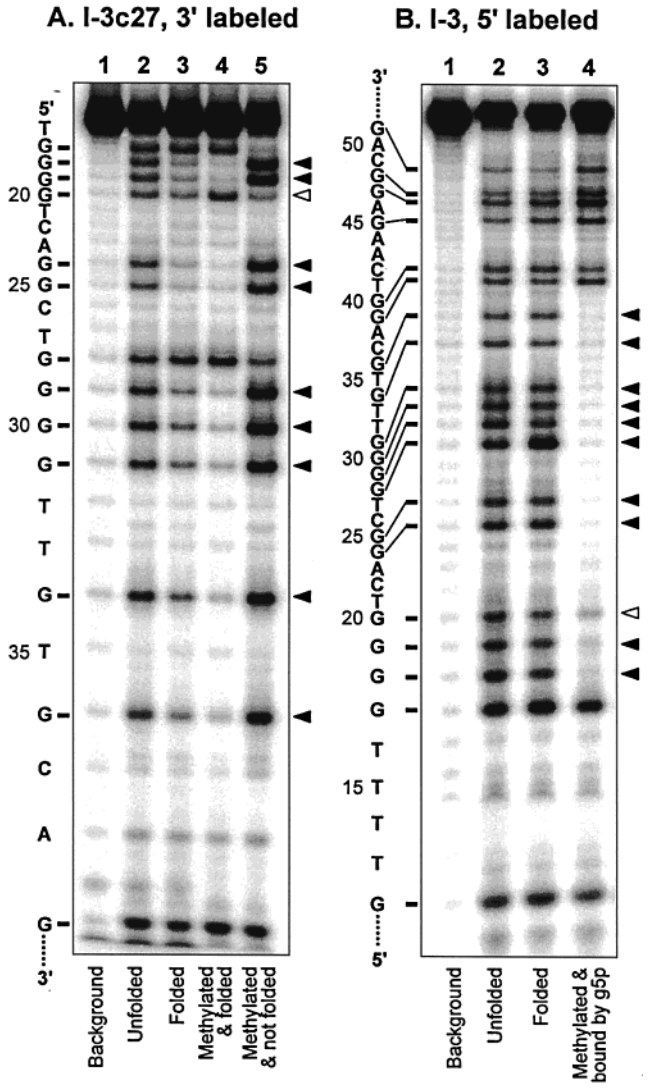

FIGURE 7: DMS methylation protection and interference. (A) DMS methylation of I-3c27. I-3c27 was ^{32}P-labeled at the 3′ end. Lane 1: background cleavage by piperidine. Lane 2: overall cleavage at guanines of the unfolded, unstructured I-3c27. Lane 3: guanines protected from methylation in the folded G-quadruplex form of I-3c27. The 3′-labeled I-3c27 was folded in 200 mM NaCl, randomly methylated by DMS, ethanol precipitated, and subjected to piperidine cleavage. Lanes 4 and 5: guanines whose methyation interfered with G-quadruplex formation. The 3′-labeled I-3c27 was randomly methylated by DMS in the absence of NaCl and then was allowed to fold in 200 mM NaCl. Following electrophoresis on native polyacrylamide gels containing 100 mM NaCl, bands corresponding to a structured form (lane 4) and an unstructured form (lane 5) were isolated. All of the samples were cleaved by piperidine, dried, and resolved on a 20% denaturing gel. (B) DMS methylation of I-3. I-3 was ^{32}P-labeled at the 5′ end. Lanes 1−3: the results of the same treatments of the 5′-labeled I-3 as for the 3′-labeled I-3c27 in corresponding lanes of panel A. Lane 4: guanines whose methyation interfered with formation of an initiation complex with g5p. The 5′-labeled I-3 was randomly methylated by DMS in the absence of NaCl and then incubated with g5p in 200 mM NaCl. Complexes were resolved by agarose gel electrophoresis. The DNA from a band corresponding to an initiation complex was isolated, subjected to piperidine cleavage, dried, and resolved on a 12% denaturing gel. The sequences of I-3c27 (panel A) and I-3 (panel B) are both numbered according to the location of the nucleotides within the I-3 sequence. Guanine bands that changed are marked with arrowheads on the right of each panel.

an extra nucleotide (T16) at its 5′ end rendered the cleavage of G17 distinguishable. Despite the fact that the GGTC-3′ end was at the bottom end of the resolving gel in the 3′-end-labeled I-3c27 experiment, it was clear that G39 in this sequence was not involved in forming a G-quartet.

Our interpretation of the data in Figure 7A is that the G's within two G_4 blocks and one G_2 block, plus those of a GTG sequence, were involved in the formation of a quadruplex structure. Moreover, since there was a thymine between two protected guanines of the G34-T35-G36 sequence, T35 also seemed to be involved in the quadruplex structure. Details of a possible structure are discussed below.

Guanines within the full I-3 sequence showed much less dramatic protection from DMS methylation in the presence of 200 mM NaCl (Figure 7B, lanes 2 and 3). Quantitation of results from triplicate experiments such as shown in lanes 2 and 3 confirmed modest protection (11−23%) at G18-G20, G29-G30, and G34. Overall, the methylation protection experiments of I-3 indicated that a G-quadruplex structure was discernible at 20 °C and that it involved most of the same guanines as in the folded I-3c27. Also, since the methylation reaction was carried out at 20 °C, there are interactions involving the 16-mer tails (Figure 3D) that could perturb or protect the core structure.

On the other hand, methylation of the same guanine N^7 positions that interfered with folding of I-3c27 clearly interfered with the formation of the g5p·I-3 initiation complex in an interference experiment. In fact, two more guanines remained unmethylated (G28) or mostly unmethylated (G20) in the g5p-bound G-quadruplex of I-3 than those that interfered with forming the I-3c27 G-quadruplex (compare numbered positions in lane 4 of panels A and B of Figure 7). This difference provided additional evidence that the structural rearrangement induced by g5p binding, as detected by the enhancement of the 260 nm CD bands of I-3 and I-3c26 during g5p binding (Figures 3A and 3B) and the kinetics experiments of Figure 6, coincided with a change in the G-quadruplex structure.

DISCUSSION

A Model for the I-3/I-3c26 Structure. Figure 8A is an illustration of a possible G-quadruplex structure for I-3c26 and for the core of I-3, based on our accumulated data. Kinetics data in this work and primer-binding and gel electrophoresis experiments in previous work (*9*) indicate that the G-quadruplex is intramolecular, and an intramolecular fold would allow the two tails of I-3 to extend from the same end of the quadruplex to provide a stretch of antiparallel strands suitable for a second stage of binding of g5p dimers. Because the CD spectrum of I-3c26 in the presence of Na^+ ion is unlike that in the presence of K^+ ion, and the K^+ form has features of a chair-type G-quadruplex in which loops connect adjacent lateral strands (*9*, *50*−*52*), we suggest that the Na^+ form of I-3c26 is a basket-type G-quadruplex having a diagonal loop, as shown in Figure 8.

The model in Figure 8A is based on the existence of three stacked quartets, which differ in base composition and are proposed to include one thymine and one adenine, as shown in Figure 8B−D. At least three lines of evidence support the involvement of three G-quartets in the structure. First, the methylation data showed that three contiguous G's of the G29-G31 stretch were protected by and interfered with

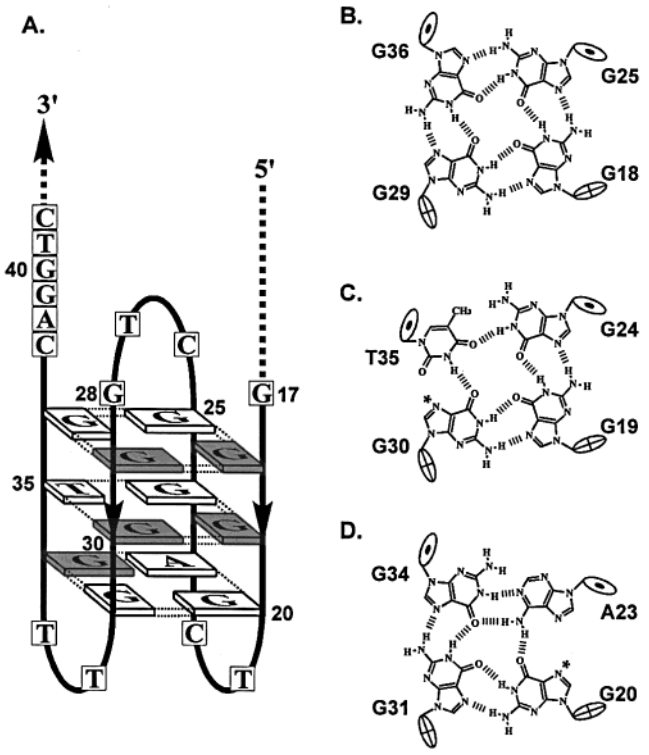

FIGURE 8: Model of a possible G-quadruplex for I-3/I-3c26. (A) A theoretical basket-type G-quadruplex fold for I-3/I-3c26 in 200 mM NaCl. Bases involved in the quartets are shown as rectangles. The dark and white rectangles, respectively, represent *syn* and *anti* conformations of glycosyl bonds. Arrows show the 5′ to 3′ strand orientation. Nucleotide numbering is with respect to I-3, and only the central 26 nucleotides of I-3 (i.e., the nucleotides of I-3c26) are shown. (B−D) Possible quartet configurations in the schematic structure in panel A. Backbone directions are indicated by the × and • symbols within the riboses, which are simplified as ovals. Gycosyl bond conformations within the quartets (*syn* with the ovals close to the base and *anti* with the ovals directed away from the base) were chosen to be similar to the arrangement in the antiparallel $d(G_3T_4G_3)$ quadruplex, including *syn* G's at the 5′ ends and *anti* G's at the 3′ ends of the quadruplex-paired strands (*39*, *40*). The inclusion of A and T bases is reasonably inferred from evidence that there are three quartets, but the hydrogen-bonding patterns (with *anti* T and A) are speculative. The N^7 positions of G20 and G30 are asterisked in (C) and (D) because these positions are not involved in hydrogen bonding according to this model. However, both of these guanines were protected or partially protected in folded I-3c26/I-3c27 (Figure 7A, lane 3).

the formation of the structure (Figure 7A). Second, CD spectra suggested that I-3c26 and G-8c26 had similar G-quadruplex structures in 200 mM NaCl (Figure 2C). G-8c26 contains four G blocks, each of which has three to four contiguous guanines that can participate in three G-quartets. Finally, the melting temperature of G-8c26 was slightly higher, by 3.5−5.7 °C, than that of I-3c26 in 10 and 200 mM NaCl, respectively (Table 1). This was consistent with the proposed I-3c26 structure, which contains one homogeneous and two less stable, mixed quartets.

Guanosines in the three quartets are shown with *syn* and *anti* glycosidic bonds arranged as in the solution structure of $d(G_3T_4G_3)$ (*39*, *40*). The G-quartets would thus have a nonalternating G-quartet polarity in the top two quartets and an alternating polarity in the bottom two quartets, possibly explaining the existence of two positive bands at 255−260 and 285−292 nm in the CD spectra of I-3 and I-3c26, as observed for $d(G_3T_4G_3)$ (*42*).

This G-quadruplex model is consistent with the methylation protection data, except that the N^7 positions of G20 and G30 are not hydrogen-bonded. The methylation of G20 did not interfere with the formation of the structure, although it was partially protected in the folded I-3c26/I-3c27 oligomers. The methylation of G30 did interfere with the formation of the quadruplex structure, and G30 was protected in the folded I-3c26/I-3c27 oligomers (Figure 7A, lanes 3 and 4). The different methylation interference results for G20 and G30 could be due to their different environments, with the space available for a methyl on an interior G30 being less than that for a methyl on G20 at the end of a quadruplex stack. The protection of G30 in its interior location in an unusual context may not be unreasonable.

Interaction of g5p with the I-3 G-Quadruplex. The N^7 positions of two guanines (G28 and G20) were required to remain unmethylated for forming the initiation complex of I-3 with g5p (Figure 7B, lane 4), but they were not required for folding the I-3c27 (or I-3c26) core into a quadruplex (Figure 7A, lane 4). The simplest assumption is that a structural change involving one or both of these guanines is related to the CD spectral changes of the free oligomers during initial g5p binding. Specifically, the guanine 255−260 nm band of I-3 and I-3c26 is nearly doubled in magnitude, while the long-wavelength positive band is changed little or not at all, during g5p binding to the core G-quadruplex structure (Figure 3A,B). Moreover, there is a kinetic block to the structural change related to the 255−260 nm band increase, which appears to return the quadruplex to its initial, transiently folded, state (Figure 6). For example, one conjecture is that an *anti*-to-*syn* conformational change of G28 would increase the nonalternating glycosyl bond pattern along the G28-G30 sequence and increase the 255−260 nm CD band, with the required glycosyl bond change providing the kinetic block. If such a change occurs in both the presence and absence of g5p, it would have to coincide with changes in base pairing that involve the N^7 position of G28. Alternatively, the N^7 positions of G20 and G28 might have to be free for stable, direct interactions with g5p and have no relationship to the perturbation of the 260 nm CD band, although this seems unlikely to us.

Contrasts and Similarities of Two g5p·G-Quadruplex Complexes. The interaction of g5p with the intramolecular G-quadruplex of I-3 stands in contrast with its interaction with the four-stranded quadruplex of 16-mer DNA strands described by Oliver et al. (*8*). The latter structure consists of four G-quartets in the center of four strands of the 16-mer DNA $d(GT_5G_4CT_4C)$. Kneale and co-workers (*8*) derived a model for the complex of this G-quadruplex and g5p, based on small-angle X-ray scattering and other data. In this model, the g5p binds only to the four six-nucleotide tails, requiring that the four strands of the intermolecular quadruplex be arranged in an unusual antiparallel fashion to accommodate the antiparallel binding sites of the g5p dimer. In the case of the g5p·I-3 complex, the antiparallel tails of the I-3 intramolecular G-quadruplex are not the first sites of binding and become saturated only after about three g5p dimers bind to the core quadruplex structure to form an initiation complex. Moreover, the affinity of g5p for I-3c26 is comparable to that for I-3 (i.e., $K\omega > 3 \times 10^7\ M^{-1}$) (*9*), suggesting that the g5p-binding affinity to I-3 is dominated by the initiation step and not by binding to the tails.

However, there may be an important similarity in the two G-quadruplex structures in that the loops and single-stranded regions of the I-3 core (Figure 8), and not just the G-quartets, are implicated in the tight binding of g5p. First, as described above, the N^7 positions of G28 and G20 both have to be free for g5p binding, suggesting a perturbation of end loop structures. Second, when the usual Na^+ cation is replaced by K^+, the structure of I-3c26 changes, probably to a chair type of G-quadruplex (*9*), to which g5p binds with lower affinity (unpublished data). That is, while the stacked G-quartets in I-3 may be directly bound by the g5p, they also appear to provide a scaffold of loops that is important for high-affinity g5p binding. If so, a similarity in the binding sites of the two G-quadruplexes may be that they both are comprised of constrained nucleotides or loops. The g5p may prefer to bind to regions of constrained single strands, as opposed to random coils, to reduce the entropic cost of binding.

There is no evidence that there are blocks of G's that could form quadruplexes in the Ff genome or Ff mRNA, so this specific class of structures may not be of direct biological relevance. It may be of more relevance that DNA or RNA structures can be found that constrain nucleotides in positions favorable for g5p binding and thereby enhance the initiation of binding under physiological salt concentrations. In addition, g5p may have higher affinity for nucleotides constrained in loops than in flexible tails. On the basis of preliminary competition experiments (not shown), we found that the affinity of g5p was >30-fold higher (in terms of strand concentration) for the I-3c26 quadruplex than for the quadruplex of the 16-mer gene 2 mRNA leader sequence (or ~8-fold higher in terms of quadruplex concentration).

G-Quadruplex Structures and Protein Binding. The most extensively studied G-quadruplex structures are usually from tandemly repeated sequences, like telomeric sequences, or from symmetric sequences, like the thrombin-binding aptamer (*52*). Thus, it is not surprising to find homogeneous G-quartets formed by these sequences, which are the prototypical G-quadruplexes. In a recent crystallographic study, the *Oxytricha* telomere DNA binding protein has been found to interact with the wide groove and connecting loops of a typical $d(G_4T_4G_4)_2$ G-quadruplex. However, these interactions are not essential for crystallizing the protein at high NaCl concentrations and may be influenced by crystal packing forces (*36*). Due to the high stability of homogeneous G-quartet structures, replacement of a guanine with another base should be acceptable as long as a majority of hydrogen bonds is retained within a minimally altered configuration. It is possible that atypical mixed G-quartets, like those that include A or T in our proposed model, coexist with typical homogeneous G-quartets and with other double chain reversal quartets (*53*, *54*). This would extend the range of sequences that can fold into quadruplex structures and provide sequence-specific recognition sites for proteins.

REFERENCES

1. Model, P., and Russel, M. (1988) in *The Bacteriophages* (Calendar, R., Ed.) pp 375−390, Plenum Press, New York.
2. Alberts, B., Frey, L., and Delius, H. (1972) *J. Mol. Biol. 68*, 139−152.
3. Gray, C. W. (1989) *J. Mol. Biol. 208*, 57−64.
4. Skinner, M. M., Zhang, H., Leschnitzer, D. H., Guan, Y., Bellamy, H., Sweet, R. M., Gray, C. W., Konings, R. N., Wang, A. H., and Terwilliger, T. C. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 2071−2075.
5. Terwilliger, T. C. (1996) *Biochemistry 35*, 16652−16664.
6. Kansy, J. W., Clack, B. A., and Gray, D. M. (1986) *J. Biomol. Struct. Dyn. 3*, 1079−1110.
7. Thompson, T. M., Mark, B. L., Gray, C. W., Terwilliger, T. C., Sreerama, N., Woody, R. W., and Gray, D. M. (1998) *Biochemistry 37*, 7463−7477.
8. Oliver, A. W., Bogdarina, I., Schroeder, E., Taylor, I. A., and Kneale, G. G. (2000) *J. Mol. Biol. 301*, 575−584.
9. Wen, J.-D., Gray, C. W., and Gray, D. M. (2001) *Biochemistry 40*, 9300−9310.
10. Michel, B., and Zinder, N. D. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 4002−4006.
11. Oliver, A. W., and Kneale, G. G. (1999) *Biochem. J. 339*, 525−531.
12. Henderson, E., Hardin, C. C., Walk, S. K., Tinoco, I., Jr., and Blackburn, E. H. (1987) *Cell 51*, 899−908.
13. Williamson, J. R., Raghuraman, M. K., and Cech, T. R. (1989) *Cell 59*, 871−880.
14. Sen, D., and Gilbert, W. (1988) *Nature 334*, 364−366.
15. Fry, M., and Loeb, L. A. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 4950−4954.
16. Simonsson, T., Pecinka, P., and Kubista, M. (1998) *Nucleic Acids Res. 26*, 1167−1172.
17. Arthanari, H., and Bolton, P. H. (2001) *Chem. Biol. 8*, 221−230.
18. Fang, G., and Cech, T. R. (1993) *Cell 74*, 875−885.
19. Giraldo, R., Suzuki, M., Chapman, L., and Rhodes, D. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 7658−7662.
20. Arimondo, P. B., Riou, J. F., Mergny, J. L., Tazi, J., Sun, J. S., Garestier, T., and Helene, C. (2000) *Nucleic Acids Res. 28*, 4832−4838.
21. Muniyappa, K., Anuradha, S., and Byers, B. (2000) *Mol. Cell. Biol. 20*, 1361−1369.
22. Chung, I. K., Mehta, V. B., Spitzner, J. R., and Muller, M. T. (1992) *Nucleic Acids Res. 20*, 1973−1977.
23. Liu, Z., and Gilbert, W. (1994) *Cell 77*, 1083−1092.
24. Sun, H., Yabuki, A., and Maizels, N. (2001) *Proc. Natl. Acad. Sci. U.S.A. 98*, 12444−12449.
25. Harrington, C., Lan, Y., and Akman, S. A. (1997) *J. Biol. Chem. 272*, 24631−24636.
26. Sun, H., Karow, J. K., Hickson, I. D., and Maizels, N. (1998) *J. Biol. Chem. 273*, 27587−27592.
27. Fry, M., and Loeb, L. A. (1999) *J. Biol. Chem. 274*, 12797−12802.
28. Sun, H., Bennett, R. J., and Maizels, N. (1999) *Nucleic Acids Res. 27*, 1978−1984.
29. Walsh, K., and Gualberto, A. (1992) *J. Biol. Chem. 267*, 13714−13718.
30. Weisman-Shomer, P., and Fry, M. (1993) *J. Biol. Chem. 268*, 3306−3312.
31. Erlitzki, R., and Fry, M. (1997) *J. Biol. Chem. 272*, 15881−15890.
32. Phillips, K., Dauter, Z., Murchie, A. I., Lilley, D. M., and Luisi, B. (1997) *J. Mol. Biol. 273*, 171−182.
33. Jin, R., Gaffney, B. L., Wang, C., Jones, R. A., and Breslauer, K. J. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 8832−8836.
34. Kang, C., Zhang, X., Ratliff, R., Moyzis, R., and Rich, A. (1992) *Nature 356*, 126−131.
35. Smith, F. W., and Feigon, J. (1992) *Nature 356*, 164−168.
36. Horvath, M. P., and Schultz, S. C. (2001) *J. Mol. Biol. 310*, 367−377.
37. Balagurumoorthy, P., Brahmachari, S. K., Mohanty, D., Bansal, M., and Sasisekharan, V. (1992) *Nucleic Acids Res. 20*, 4061−4067.
38. Lu, M., Guo, Q., and Kallenbach, N. R. (1993) *Biochemistry 32*, 598−601.
39. Smith, F. W., Lau, F. W., and Feigon, J. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 10546−10550.
40. Keniry, M. A., Strahan, G. D., Owen, E. A., and Shafer, R. H. (1995) *Eur. J. Biochem. 233*, 631−643.
41. Williamson, J. R. (1994) *Annu. Rev. Biophys. Biomol. Struct. 23*, 703−730.
42. Scaria, P. V., Shire, S. J., and Shafer, R. H. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 10336−10340.
43. Laporte, L., Benevides, J. M., and Thomas, G. J., Jr. (1999) *Biochemistry 38*, 582−588.
44. Catasti, P., Chen, X., Moyzis, R. K., Bradbury, E. M., and Gupta, G. (1996) *J. Mol. Biol. 264*, 534−545.
45. Hardin, C. C., Henderson, E., Watson, T., and Prosser, J. K. (1991) *Biochemistry 30*, 4460−4472.

46. Mou, T.-C., Gray, C. W., and Gray, D. M. (1999) *Biophys. J. 76*, 1537−1551.
47. Hashem, G. M., Pham, L., Vaughan, M. R., and Gray, D. M. (1998) *Biochemistry 37*, 61−72.
48. Maxam, A. M., and Gilbert, W. (1980) *Methods Enzymol. 65*, 499−560.
49. Guo, Q., Lu, M., and Kallenbach, N. R. (1992) *J. Biol. Chem. 267*, 15293−15300.
50. Wang, K. Y., McCurdy, S., Shea, R. G., Swaminathan, S., and Bolton, P. H. (1993) *Biochemistry 32*, 1899−1904.
51. Smirnov, I., and Shafer, R. H. (2000) *Biochemistry 39*, 1462−1468.
52. Macaya, R. F., Schultze, P., Smith, F. W., Roe, J. A., and Feigon, J. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 3745−3749.
53. Kuryavyi, V., Majumdar, A., Shallop, A., Chernichenko, N., Skripkin, E., Jones, R., and Patel, D. J. (2001) *J. Mol. Biol. 310*, 181−194.
54. Wang, Y., and Patel, D. J. (1994) *Structure 2*, 1141−1156.

BI020276E