Ff Gene 5 Single-Stranded DNA-Binding Protein Assembles on Nucleotides Constrained by a DNA Hairpin†

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 July 24, 2003; Revised Manuscript Received December 17, 2003

ABSTRACT: The gene 5 protein (g5p) encoded by filamentous Ff phages is an ssDNA-binding protein, which binds to and sequesters the nascent ssDNA phage genome in the process of phage morphogenesis. The g5p also binds with high affinity to DNA and RNA sequences that form G-quadruplex structures. However, sequences that would form G-quadruplexes are absent in single copies of the phage genome. Using SELEX (*s*ystematic *e*volution of *l*igands by *ex*ponential enrichment), we have now identified a family of DNA hairpin structures to which g5p binds with high affinity. After eight rounds of selection from a library of 58-mers, 26 of 35 sequences of this family contained two regions of complete or partial complementarity. This family of DNA hairpins is represented by the sequence: 5′-d(C*GGGATCC*AA-CGTTTTCACCAGATCTACCTCCTCG*GGATCCC*AAGAGGCAGAATTCGC)-3′ (named U-4), where complementary regions are italicized or underlined. Diethyl pyrocarbonate modification, UV-melting profiles, and *BamH* I digestion experiments revealed that the italicized sequences form an intramolecular hairpin, and the underlined sequences form intermolecular base pairs so that a dimer exists at higher oligomer concentrations. Gel shift assays and end boundary experiments demonstrated that g5p assembles on the hairpin of U-4 to give a discrete, intermediate complex prior to saturation of the oligomer at high g5p concentrations. Thus, biologically relevant sequences at which g5p initiates assembly might be typified better by DNA hairpins than by G-quadruplexes. Moreover, the finding that hairpins of U-4 can dimerize emphasizes the unexpected nature of sequence-dependent structures that can be recognized by the g5p ssDNA-binding protein.

The genome of the Ff viruses[1] is a circular ssDNA (plus strand) comprising 11 tightly packed genes and a 508-nucleotide intergenic region (IG) (*1*, *2*). The IG contains five functional hairpin structures, denoted hairpins A−E (*3*), which include some regulatory elements, such as the morphogenetic signal for packaging (hairpin A), the minus (complementary) strand origin (involving hairpins B and C), and the plus strand origin on hairpin D (*4*).

The viral gene 5 encodes an ssDNA-binding protein, the g5p, which sequesters the single-stranded genome into a superhelical complex that is a precursor of virus assembly at a late stage of infection (*5*, *6*). The g5p exists as a homodimer with a 2-fold rotational axis of symmetry so that its binding clefts are positioned to bind two antiparallel ssDNA strands (*7*−*9*). 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 the g5p binds to genomic or purine-rich DNA at P/N ratios less than 0.25 (*10*, *11*). The binding affinity ($K\omega$) is the result of an intrinsic binding constant (K) multiplied by a cooperativity factor (ω). The values of ω are estimated to be in the range of 500−5000 (*9*, *12*, *13*). Thus, g5p saturates ssDNA with a high cooperativity of binding.

Although the g5p can bind to any nucleic acid sequence, it has been shown that the binding affinity for synthetic ssDNA depends on the base composition (*12*, *13*). For example, the $K\omega$ for dC_{48} is about 140-fold higher than for dA_{48} at 0.2 M NaCl. The sequence preferences of g5p binding are biologically important in that the g5p regulates mRNA translation of the viral genes 1, 2, 3, 5, and 10 via specific leader sequences of the mRNA (*14*, *15*). Among these sequences, the gene 2 mRNA leader sequence has been most studied. A 16-mer sequence, $r(GU_5G_4CU_4C)$, that is within the gene 2 mRNA leader region is required in vivo for translational repression by the g5p (*14*). Mutations in this sequence that abolish gene 2 translational repression in vivo show reduced g5p-binding affinity in vitro (*16*). These results suggest that the g5p specifically binds to this sequence in performing its regulatory function. Further studies show that

† Support was provided by Grant AT-503 from the Robert A. Welch Foundation and in part by Grant 009741-0021-1999 from the Texas Advanced Technology Program.

* Corresponding author. Phone: (972) 883-2513. Fax: (972) 883-2409. E-mail: dongray@utdallas.edu.

[1] Abbreviations: binding buffer, 200 mM NaCl, 1 mM EDTA, and 10 mM Tris-HCl, pH 7.4; DMS, dimethyl sulfate; DEPC, diethyl pyrocarbonate; EMSA, electrophoretic mobility shift assay; Ff viruses, three closely related filamentous viruses f1, fd, and M13 that specifically infect F^+ strains of *Escherichia coli*; g5p, gene 5 protein; IG, intergenic region; P/N, the [protein monomer]/[nucleotide] molar ratio; SDS−PAGE, sodium dodecyl sulfate−polyacrylamide gel electrophoresis; SELEX, *s*ystematic *e*volution of *l*igands by *ex*ponential enrichment; ssDNA, single-stranded DNA; TBE buffer, 90 mM Tris-borate, pH 8.3, 2 mM EDTA; TE buffer, 10 mM Tris-HCl, pH 7.4, 1 mM EDTA; T_m, melting temperature.

10.1021/bi030177g CCC: $27.50
Published on Web 02/12/2004

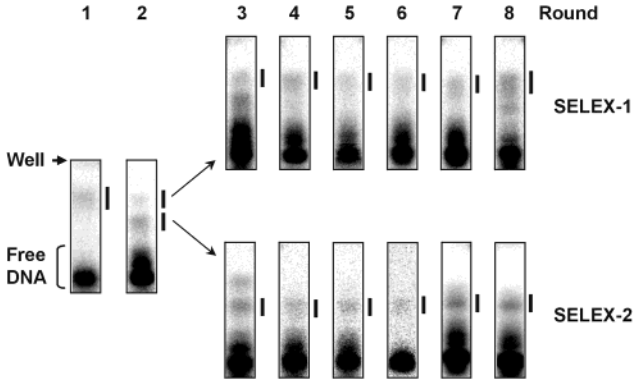

FIGURE 1: Images of gels during SELEX selections. Phosphor images are shown of the bands in agarose gels used to separate g5p·DNA complexes from free DNA during the selection steps of SELEX-1 and -2. Positions of the loading well and free DNA are indicated. Positions of complexes isolated for the subsequent rounds of SELEX are indicated with a black bar to the right of the image. The upper band of saturated complex in round 2 was recovered and followed through subsequent rounds for SELEX-1 (top images). The lower band of intermediate unsaturated complex was recovered and followed through subsequent rounds for SELEX-2 (bottom images). Complexes isolated from the eighth round of each set of SELEX selections were cloned and sequenced. Selection ratios (complexed DNA/total DNA) at individual steps ranged from 0.005 to 0.05 for SELEX-1 and from 0.018 to 0.037 for SELEX-2. The multiple bands that were observed in the range of free DNA were probably from DNA multimers (of which the U-4 dimer is an example; see Figure 3A).

the 16-mer operator, in the form of either RNA or DNA, can form a four-stranded G-tetraplex structure through the G_4 blocks at the center of the sequence (*17, 18*). This G-quadruplex structure was proposed by Kneale and coworkers to hold adjacent single-stranded regions in position for binding of the g5p (*17, 18*). To be biologically relevant, such a structure would have to form by intermolecular pairing of four mRNA molecules.

Use of the SELEX technique (*19, 20*) has allowed us to identify in a more comprehensive fashion the preferred DNA binding sites for the g5p (in 200 mM NaCl, pH 7.4, 37 °C). In the first set of experiments, we isolated DNA from a g5p-saturated complex on agarose gels (SELEX-1, see Figure 1). A family of G-rich sequences was identified, and a study of the dominant sequence, I-3, showed that it formed an intramolecular G-quadruplex structure with two extended tails (*21, 22*). In this case, the g5p bound to I-3 in two stages, first to the G-quadruplex core and then to the flanking tails. Although the binding of g5p to this G-quadruplex differs from that to the G-quadruplex described by Kneale and coworkers, in both cases nucleotides are constrained in positions favorable for binding (*22*). That is, high affinity sites for g5p binding may consist of appropriately constrained regions of ssDNA to reduce the entropic cost of binding. Notwithstanding this insight, there is no evidence that blocks of Gs form quadruplexes within the Ff genome or Ff mRNA.

We now have evidence for another family of high-affinity binding sites that may be more relevant to the in vivo binding of g5p. During SELEX-1 selection, DNA was isolated from a g5p-saturated complex. However, an intermediate, unsaturated complex appeared during the second round. Therefore, we performed another set of SELEX experiments (termed SELEX-2) on the intermediate complex from the second round. The unprecedented finding is that g5p binds to DNA hairpins that have the additional ability to dimerize.

EXPERIMENTAL PROCEDURES

Protein Purification. The g5p was purified to 99% homogeneity from *Escherichia coli* transformed with an Ff gene 5-containing plasmid, as described previously (*11, 13*).

SELEX. A 58-mer ssDNA library was used for the initial SELEX selection. The library, named PV-58, was synthesized with the following sequence: 5′-d(CGGGATCCAACGTTTT-N_{26}-AAGAGGCAGAATTCGC)-3′ (Oligos Etc., Wilsonville, OR). A, G, C, and T were randomly incorporated in the central 26 nucleotides (N_{26}). The SELEX procedure is detailed in previous work (*21*). In brief, to approximate physiologically relevant binding conditions, the g5p-DNA binding reaction was performed at 37 °C in binding buffer (200 mM NaCl, 1 mM EDTA, and 10 mM Tris-HCl, pH 7.4), and g5p-bound DNA was separated on agarose gels. In that work, a saturated g5p·DNA complex was isolated and subjected to seven further rounds of selection [termed SELEX-1 (*21*)]. In the present work, we isolated the DNA from an intermediate, unsaturated complex (termed SELEX-2), except for the first round of selection, in which only a saturated complex was formed. As in SELEX-1, selected DNA was eluted from agarose gels and amplified by PCR. A reverse primer with a biotin tag was used during the amplification such that the target strand could be recovered by binding the PCR products to streptavidin beads. After eight rounds of selection, the DNA was cloned and sequenced.

DNA Oligomers. Specific DNA sequences were purchased from Oligos Etc. or Midland (Midland, TX). The 58-mer oligomer (named U-4) was 5′-d(CGGGATCCAACGTTTT-CACCAGATCTACCTCCTCGGGATCCC-AAGAGGCAGAATTCGC)-3′; two pairs of complementary regions are underlined or double-underlined. U-4t42 was a sequence having the first 42 nucleotides of U-4: 5′-d(CGGGATCC-AACGTTTT-CACCAGATCTACCTCCTCGGGATCCC)-3′. The hp42 sequence was identical to U-4t42 except for four mutated nucleotides (shown in lower case) in the complementary regions: 5′-d(CGGtcgtCAACGTTTT-CACCA-GATCTACCTCCTCGGacgaCC)-3′; complementarity was maintained in the underlined regions. Ss42 was another mutant U-4t42 sequence, but the complementarity was reduced to five nucleotides: 5′-d(CGGctTCCAACGTTTT-CACCAGATCTACCTCCTCGGatTCCC)-3′; mutated nucleotides are shown in lower case. 16-mer primers were 5′-d(AAAACGTTGGATCCCG)-3′ (named C5′P) and 5′-d(GCGAATTCTGCCTCTT)-3′ (named C3′P), which were complementary to the 5′- and 3′-ends of U-4, respectively. Concentrations of oligomers were determined by absorption measurements and nearest-neighbor formulas as described by Gray et al. (*23*), and concentrations reported throughout this work are per strand.

EMSA and Competition Experiments. EMSA for g5p-DNA binding and competition reactions was performed on 2.5% agarose gels as described in previous work (*21*). In determining the relative binding affinity, 1 μM U-4t42 was ^{32}P-labeled at the 5′-end by the T4 polynucleotide kinase (Promega, Madison, WI) and was mixed with serial concentrations of unlabeled DNA oligomers as competitors. The g5p was

added to 6 μM, and the reaction was performed in binding buffer at 37 °C for 15 min, followed by electrophoresis on agarose gels as in EMSA. The relative g5p binding affinity was determined to be the concentration ratio of competitor/U-4t42 at which 50% of labeled U-4t42 was dissociated.

Stoichiometry of Protein and DNA in g5p·U-4 Complexes. The stoichiometric ratio of protein to DNA in g5p·U-4 complexes was determined as described in previous work (*21*). In brief, ^{32}P-labeled U-4 was incubated with g5p in binding buffer, and the mixture was separated on agarose gels. The intermediate and saturated complexes were isolated and subjected to electrophoresis based on SDS−PAGE (*24*). A series of standard amounts of g5p and ^{32}P-labeled U-4 were run in parallel on the same gel. The protein was quantitated by staining the gel with SYPRO Red (BMA, Rockland, ME) (*25*), and then the gel was dried and exposed to a storage phosphor screen (Molecular Dynamics, Sunnyvale, CA) for DNA quantitation.

Primer-Annealing Experiments. ^{32}P-labeled U-4 was prepared at concentrations of 10, 1, 0.1, and 0.01 μM with equal radioactivity. Various amounts of primer oligonucleotides (C5′P or C3′P) were added such that the strand ratio (primer/U-4) was 1.5 at U-4 concentrations of 10, 1, and 0.1 μM and was 5 at the U-4 concentration of 0.01 μM. Mixtures were prepared in 10 μL of binding buffer. Samples were heated at 95 °C for 3 min, slowly cooled at room temperature for 30 min, and resolved on 12% native polyacrylamide gels in TBE buffer (90 mM Tris-borate, pH 8.3, 2 mM EDTA) containing 100 mM NaCl. The gels were dried and exposed to a storage phosphor screen.

UV Melting. UV-melting profiles of oligonucleotides were obtained at 260 nm on 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 before the acquisition of absorbance data. The melting temperature (T_m) was defined as the local maximum of the first derivative of the melting profile.

DEPC Modification Protection. U-4 was labeled at the 3′-end with T4 RNA ligase (Promega) and [5′-^{32}P]-pCp (ICN, Irvine, CA). The 10 μM labeled U-4 was mixed with 15 μM C5′P or C3′P in a total volume of 20 μL of binding buffer, heated at 95 °C for 3 min, and slowly cooled at room temperature for 30 min. One microliter of DEPC (Sigma, St. Louis, MO) was added and incubated at room temperature for 1 h, followed by ethanol precipitation to stop the reaction. The modified DNA was cleaved by piperidine (Aldrich, Milwaukee, WI) as per Maxam and Gilbert (*26*) except for a reduction in cleavage time to 15 min at 90 °C. Sequence fragments were separated on 12% sequencing gels in TBE buffer.

DEPC/DMS Modification Interference. To determine the interference of DEPC or DMS modification on the formation of the g5p·U-4 intermediate complex, 5′-^{32}P-labeled U-4 was prepared in TE buffer (10 mM Tris-HCl, pH 7.4, 1 mM EDTA) and randomly modified at 37 °C by DEPC for 30 min or by DMS (Aldrich) for 15 min. DNA was extracted with phenol/chloroform, precipitated in ethanol, and resuspended in TE buffer. Modified U-4 (2 μM) was allowed to fold in binding buffer for 15−30 min at room temperature and then was titrated with g5p at 37 °C for 15 min. The resulting complexes were resolved on agarose gels as for the EMSA experiments. The bands corresponding to the intermediate and saturated complexes were isolated, and the DNAs were extracted, cleaved by piperidine, and applied to 20% polyacrylamide sequencing gels in TBE buffer.

Restriction Digestion of BamH I. 5′-^{32}P-labeled U-4 was prepared at concentrations of 10, 1, 0.1, and 0.01 μM with equal radioactivity. The U-4 samples were heated at 95 °C for 3 min and slowly cooled at room temperature for 30 min, in a buffer containing 6 mM Tris-HCl, pH 7.5, 150 mM NaCl, 6 mM $MgCl_2$, and 1 mM DTT (dithiothreitol). *BamH* I (Promega) was added to give a final concentration of 0.5 U/μL. The mixtures were incubated at 37 °C for 30 min and resolved on 20% polyacrylamide sequencing gels in TBE buffer.

End Boundaries of U-4 for the Intermediate Complex. To determine the 5′- and 3′-end boundaries of the U-4 sequence needed to form the intermediate complex with g5p, U-4 was ^{32}P-labeled at the 3′- and 5′-ends, respectively. Since some nucleases (like DNase I and nuclease S1) cleaved U-4 only at limited sites in this case (not shown), a more diverse pool of U-4 fragments was obtained by combining the fragments generated by DNase I (Sigma) and by piperidine following DEPC/DMS modification (see above). The U-4 fragments (∼0.18 or ∼1 mM in nucleotide, corresponding to ∼3 or ∼17 μM of the full-length U-4 sequence, respectively) were prepared in binding buffer, titrated with g5p at 37 °C for 15 min, and resolved on agarose gels as for the EMSA experiments. The bands corresponding to the intermediate and saturated complexes were isolated, and the DNA was extracted and applied to 8 or 12% polyacrylamide sequencing gels.

RESULTS

SELEX Selection from an Intermediate Complex. A 58-mer ssDNA library, PV-58, was used for initial SELEX selection for g5p binding. The reaction was performed under physiologically relevant conditions: pH 7.4, 200 mM NaCl, and 37 °C. After eight rounds of selection, 35 of the DNA sequences from SELEX-2 (Figure 1) were cloned and sequenced. Thirty-four of these sequences were grouped into two families, leaving one unclassified sequence (Table 1). The variable region of sequences in family II was G-rich, and all but one had four blocks of G_{2-4}, a feature similar to that of previously selected sequences that can form G-quadruplexes (*21*). This family of sequences will not be considered further here. Our focus is on the major family I group of sequences, which included 26 of the 35 sequences. Family I sequences had three common aspects. First, a sequence centered at GATCCC (referred to as the α segment, see Table 1) was conserved and was complementary to nucleotides in the 5′-end constant region (the α′ segment). The α segment was at or close to the downstream end of the variable region. Second, a pyrimidine-rich sequence (the β segment) had partial complementarity to purine-rich nucleotides in the 3′-end constant region (the β' segment). The potential β:β' pairing did not contain more than five continuous base pairs, and the β segment was always just a few bases upstream of the α segment. Third, for most sequences, the α segment included a palindrome (GGATCC or GATC) and was located at about the midpoint of the β

Table 1: Thirty-Five Sequences after Eight Rounds of Selection[a]

	α′	β	α	β′
Family I				
U-4	cgggatccaacgttttCACCAGATCTACCTCCTCGGGATCCCaagaggcagaattcgc			
U-7	cgggatccaacgttttATGCGCCCCTACTCTGCTAGGATCCCaagaggcagaattcgc			
T-7 (2)[b]	cgggatccaacgttttGCCCACACTCCCGCTTCCAGGATCCCaagaggcagaattcgc			
V-10	cgggatccaacgttttCATCACACCATTCCGCTGGGGATCCCaagaggcagaattcgc			
V-14	cgggatccaacgttttAGGGATGCACTCCTCCTAGGGATCCCaagaggcagaattcgc			
V-1	cgggatccaacgttttCCCACACTCACTGGCATCGGATCCCCaagaggcagaattcgc			
U-6	cgggatccaacgttttCACCGGCACCTTCATTGGGGATCCCTTaagaggcagaattcgc			
T-2	cgggatcaacgttttGCGCACCTTGTCCACCTTCTGATCCCaagaggcagaattcgc			
T-5	cgggatcaacgttttCACTCGCGGCCATGCTTATTGATCCCaagaggcagaattcgc			
U-11	cgggatcaacgttttCACTACACTCTCGCCACTGTGATCCCaagaggcagaattcgc			
V-15	cgggatcaacgttttCCACTGCGCCCTCCCTTCTTGATCCCaagaggcagaattcgc			
V-17	cgggatcaacgttttCACAGGACTCACCCTTGCATGATCCCaagaggcagaattcgc			
U-10	cgggatcaacgttttCCGCAGCTACATCCTCATGATCCCTTaagaggcagaattcgc			
V-12	cgggatcaacgttttCGCCCACACTCCCTCCTTGATCCCCAaagaggcagaattcgc			
V-16	cgggatcaacgttttCACACCCTGCACTCCTGTGATCCCTGaagaggcagaattcgc			
V-5	cgggatcaacgttttCCACTCACGCTCTCTCGATCCCACATaagaggcagaattcgc			
U-2	cggatccaacgttttCACGCTCACTGTCCCTTACGGATCCTaagaggcagaattcgc			
U-1	cggatccaacgttttCACAGTCCTCCTCCCCAGGATCGCGTaagaggcagaattcgc			
T-3	cgggatccaacgttttGCATACTCACTCCCTACATCCCGTGCaagaggcagaattcgc			
V-18	cgggatccaacgttttCACATTCACTCTGCTCCATCCCGTGTaagaggcagaattcgc			
U-5	cgggatccaacgttttCGGGCACGTGTCCTTGTCTTCCCGCCaagaggcagaattcgc			
U-8	cgggatccaacgttttCCACACTCTACTCCCTGGGGTTCCCGaagaggcagaattcgc			
V-2	cgggatccaacgttttCATCCACTCCTTCCATCGGGTCCCGCaagaggcagaattcgc			
T-1	cgggatccaacgttttCACACTCACCGCATCCATTGGTCCCGaagaggcagaattcgc			
U-9	cgggatccaacgttttGCCGCCACTTCCTCTGGTTGGTCCCaagaggcagaattcgc			
Family II				
U-3 (3)[b]	cgggatccaacgttttATCAGGGGCGGTCGGGGCTCGGGCGTaagaggcagaattcgc			
T-4	cgggatccaacgttttGAGGGGCGGTCGGGGCTCGGGCATTGaagaggcagaattcgc			
T-6	cgggatccaacgttttCAGGGGCGGCCGGGGCTCGGGCTTCTaagaggcagaattcgc			
V-7	cgggatccaacgttttCAAGGGGTGGTCGGGGCTCGGGCAGTaagaggcagaattcgc			
V-9	cgggatccaacgttttGGGGTGGCAAGGGGTTTCCGGTCGCTaagaggcagaattcgc			
T-8	cgggatccaacgttttAAGGAGGGGACAACGGCGTGCGTGCCaagaggcagaattcgc			
Other				
V-8	cgggatccaacgttttCCACGCCCCCTCCTTTGCTACGTTGGaagaggcagaattcgc			

[a] Complementary sequences are boxed and underlined; blocks of Gs are double underlined. The 5′- and 3′-end constant sequences are shown in lower case, and sequences derived from the random region are shown in upper case. [b] The figures in parentheses are the numbers of clones with identical sequences.

and β′ segments. Therefore, the complementary and symmetrical features suggested that family I sequences could form a structure that differed from the G-quadruplex structures of family II. Simultaneous pairing of both α: α′ and β: β′ in a single molecule would result in a pseudoknot structure. Possible ssDNA pseudoknots that bind to HIV-1 reverse transcriptase have also been identified by SELEX (*27*), although most examples of possible pseudoknotted ligands have been selected from RNA libraries (*28*). In the present case, our results showed that a structure much different from an ssDNA pseudoknot was actually formed.

Two Stages of Binding of g5p to U-4. We chose a representative sequence, U-4, from family I for detailed study. U-4 had a typical α segment and was among those with maximal β: β′ complementarity, the structural implication of which we wished to ascertain. Titration of U-4 with g5p was performed in 200 mM NaCl, and the resulting complexes were resolved on agarose gels. Figure 2A shows the results of this electrophoretic mobility shift assay (EMSA). The g5p bound to U-4 in two stages: an intermediate complex was formed early in the titration, and this was followed by formation of a saturated complex. The intermediate and saturated complexes contained about 10 and 15 g5p monomers, respectively, per U-4 strand (Table 2), confirming that the intermediate complex was not saturated. The two-stage binding was reminiscent of EMSA experiments with I-3 (a SELEX-1 sequence that forms a G-quadruplex), as shown in Figure 2B for comparison (*21*). However, two major differences were evident in the EMSA data. First, the mobility of the intermediate complex of g5p·U-4 was more retarded than that of the initiation complex of g5p·I-3. [The former was formed with a stoichiometry of about 10 protein monomers per strand, whereas the latter contained only about six monomers per strand (*21*).] Second, the saturated complex of g5p·U-4 appeared only at high P/N ratios, so that the population of the intermediate complex reached a peak of about 80% of the DNA at P/N = 0.19 (lane 7, Figure 2A,C). The g5p·I-3 saturated complex readily

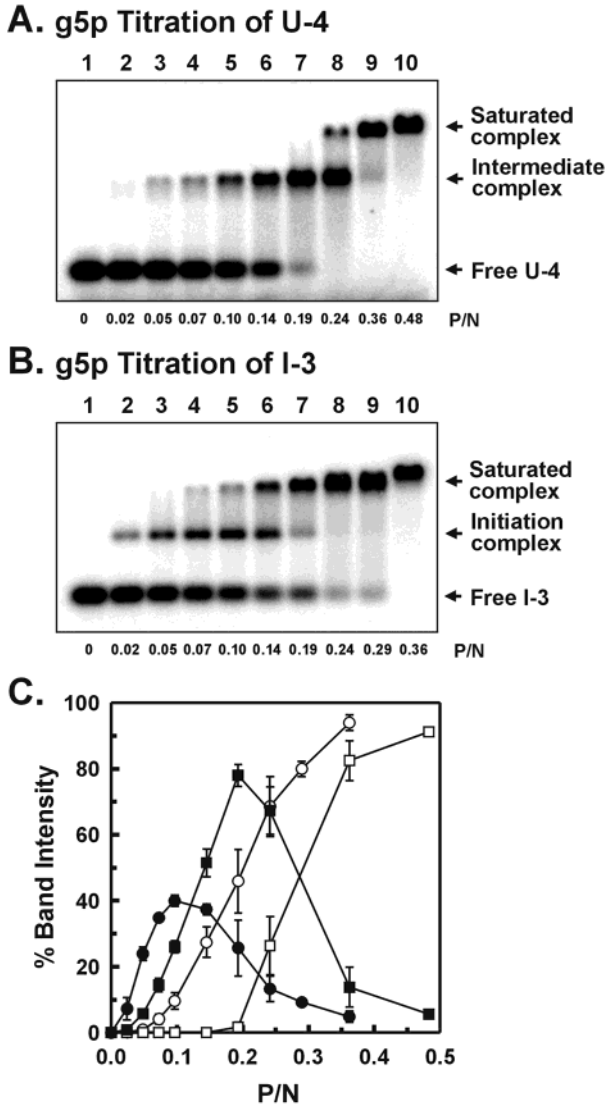

FIGURE 2: EMSA of g5p titrations of U-4 and I-3. ^{32}P-labeled U-4 (A) or I-3 (B) at 1 μM was titrated with increasing concentrations of g5p at 200 mM NaCl, 37 °C. Samples were subjected to electrophoresis on 2.5% agarose gels. P/N ([protein monomer]/[nucleotide]) molar ratios are shown at the bottom of the panel. The protein concentration at a P/N ratio of 0.25 was theoretically sufficient to saturate the DNA in the $n = 4$ binding mode. (C) Quantitation of bands from EMSA gels. Percentages of the bands for each P/N ratio in panels A and B are shown for the saturated (□) and intermediate (■) complexes from U-4 and for the saturated (○) and initiation (●) complexes from I-3. Data were averages from three independent titrations, and error bars are standard deviations. In panels A and B of this figure, and in Figures 3, 4, and 6−9, phosphor images are shown of representative gels after fixing and drying.

Table 2: Protein/DNA Stoichiometries[a] and Corresponding P/N Ratios of the G5p·U-4 Complexes

	g5p monomer/U-4 strand[b]	P/N ratio
saturated complex	15.1 ± 1.8	~0.25
intermediate complex	9.7 ± 1.7	~0.17

[a] See Experimental Procedures for details. [b] Data are shown as mean ± standard deviation from at least three measurements.

formed at relatively low P/N ratios (P/N ≥ 0.07, Figure 2B), and the intermediate complex with I-3 did not exceed 40% of the DNA (Figure 2C). [The intermediate complex with I-3 has been termed an initiation complex because the g5p binds first to the core G-quadruplex structure and then to the 3′- and 5′-tails (*21*, *22*).] These differences resulted from the differences in selection (U-4 and I-3 were respectively selected from sequences that formed intermediate and saturated complexes) and from the differences in the g5p/strand stoichiometries of the intermediate complexes.

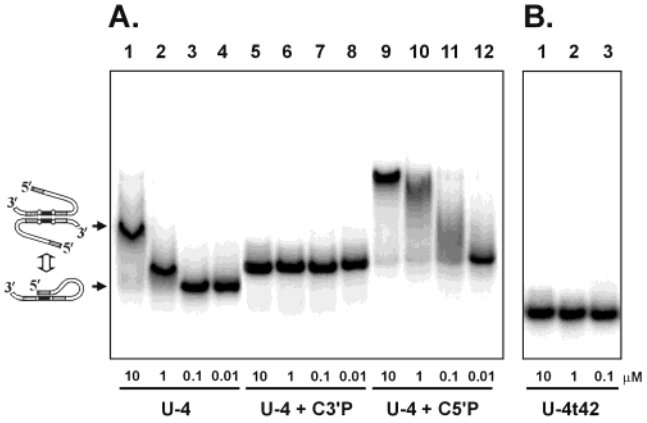

FIGURE 3: Effects of concentration and primer binding on the gel mobility of U-4 and U-4t42 DNA sequences. (A) 10-fold serial dilutions of U-4 were prepared in 200 mM NaCl with a fixed amount of ^{32}P-labeled U-4 in each dilution (lanes 1−4). Excess C3′P primer (lanes 5−8) or excess C5′P primer (lanes 9−12) was added such that the molar ratio of primer/U-4 was at least 1.5 (see Experimental Procedures). (B) U-4t42 was prepared and treated as in lanes 1−3 of panel A. All samples were heated and cooled before applying to 12% native polyacrylamide gels in TBE buffer containing 100 mM NaCl. Final concentrations of the oligomers are indicated at the bottom of the panel. (Throughout this paper, oligomer concentrations are given per strand.) Schematic drawings of the proposed U-4 dimer and monomer structures (see text) are shown to the left of their band positions.

The previous titration experiments suggested that the g5p-saturated complexes of I-3 and U-4 were formed via binding of g5p to preformed intermediate complexes. To test this hypothesis in the case of I-3, excess cold I-3 was added to preformed, saturated g5p·I-3 complexes in which I-3 was ^{32}P-labeled. The reaction mixtures were incubated for various times and then resolved on an agarose gel. The results showed that the saturated complexes gradually dissociated into intermediate complexes plus free DNA in the presence of the excess cold I-3 (data not shown). Thus, in the case of I-3, and we presume in the case of U-4, the saturated complexes most likely formed via addition of g5p to intermediate, precursor complexes.

Secondary Structures of U-4. The formation of a g5p·U-4 intermediate complex suggested that U-4 existed in a structured form (*22*). The presence of two sets of complementary sequences implied that the U-4 structure involved Watson−Crick base pairing and not G-quartets, as was the case for I-3. To investigate the U-4 structure, we first used native gel electrophoresis (Figure 3) and absorbance melting profiles (summarized in Table 3). On native polyacrylamide gels, U-4 migrated as discrete bands with concentration-dependent mobility (Figure 3A, lanes 1−4). The mobility gradually increased with decreasing DNA concentration until the concentration was reduced to about 0.1 μM. The concentration-dependent gel mobility suggested that U-4 existed in at least two fast-equilibrating states, including a dimer (or multimer) and a monomer, with the latter being predominant at or below 0.1 μM. This point of view was

Table 3: Melting Temperatures (T_m) and Percent Hyperchromicities (%H) of the Dominant Transitions in the Melting Profiles of U-4 and U-4t42

	first (minor) transition (interstrand interaction)		second (major) transition (intrastrand or bound primer interactions)	
	T_m (°C)[a]	%H at T_m[a,b]	T_m (°C)[a]	%H at T_m[a,b]
U-4				
0.2 μM	30−34[c]	∼7[c]	52.5 ± 0.7	14.3 ± 0.5
1 μM	33.4 ± 0.7	8.1 ± 0.4	52.3 ± 0.6	16.4 ± 1.3
10 μM	38.0 ± 1.1	8.8 ± 0.5	53.1 ± 1.1	15.7 ± 0.4
1 μM + C3′P[d]	n.d.	n.d.	59.0 ± 0.8	15.8 ± 0.7
1 μM + C5′P[d]	42.2 ± 0.6	8.7 ± 1.1	60.7 ± 0.3	20.8 ± 1.2
U-4t42				
1 μM	n.d.	n.d.	51.1 ± 0.4	12.2 ± 0.3
1 μM + C5′P[d]	n.d.	n.d.	60.5 ± 0.5	16.7 ± 0.5

[a] Data are shown as mean ± standard deviation from at least three measurements. The melting was performed in a buffer of 10 mM Tris-HCl, the pH of which changed from 7.5 (at 20 °C) to 6.3 (at 75 °C). n.d., not detected. [b] Percent hyperchromicity values were determined as the percent increase in absorbance at 260 nm at the T_m, relative to the absorbance at the lowest temperature (∼5 °C). [c] The first transition at 0.2 μM U-4 was not obvious or disappeared in some measurements. The T_m and %H values shown here are approximate. [d] The concentration of C5′P or C3′P was 1.5 μM.

supported by absorbance melting experiments. U-4 showed two transitions in melting profiles (Table 3). The first (minor) transition was 38.0 °C at a U-4 concentration of 10 μM, decreased to 33.4 °C at 1 μM, and became obscure or disappeared (in some measurements) at 0.2 μM. Thus, given the gel electrophoresis data, the first transition was apparently related to interstrand interactions, which were minor or missing at low concentrations (0.2 μM or less). The second (major) transition, in contrast with the first one, was consistent in the T_m (52−53 °C, Table 3) for all the measured concentrations and likely corresponded to intrastrand structures.

Interstrand interactions of U-4 were presumably through base pairing of complementary sequences, including 5′- and/or 3′-end constant regions (see Table 1). The first evidence that interstrand interactions involved the 3′-end came from data with U-4t42, a truncated sequence of U-4 lacking the 16-mer 3′-end constant segment. U-4t42 migrated as a uniform species on native polyacrylamide gels, regardless of concentration (Figure 3B). To further test the involvement of the two ends, we used 16-mer oligonucleotides, C5′P and C3′P, which were complementary to the 5′- and 3′-end constant sequences of U-4, respectively. As shown in Figure 3A (lanes 5−8), hybridization of C3′P shifted U-4 at all concentrations to a single position on the gel, consistent with one C3′P being hybridized to one U-4 strand. Melting profiles also showed that the first transition disappeared when the 3′-end of U-4 was deleted (i.e., U-4t42) or was hybridized with C3′P (Table 3). Thus, the 3′-end constant region of U-4 was required for the interstrand interaction, and blocking the 3′-end drove the equilibrium of U-4 structures to a monomer state. In contrast, the second T_m persisted for U-4t42 (Table 3), and it was present over a wide range of U-4 concentrations (0.2−10 M). Therefore, the first transition at 33−38 °C was assigned to interstrand interactions involving the 3′-end, and the second transition at 51−53 °C was attributed to intrastrand interactions.

Intra- and intermolecular interactions involving the 5′-end constant region of U-4 were tested by hybridization with

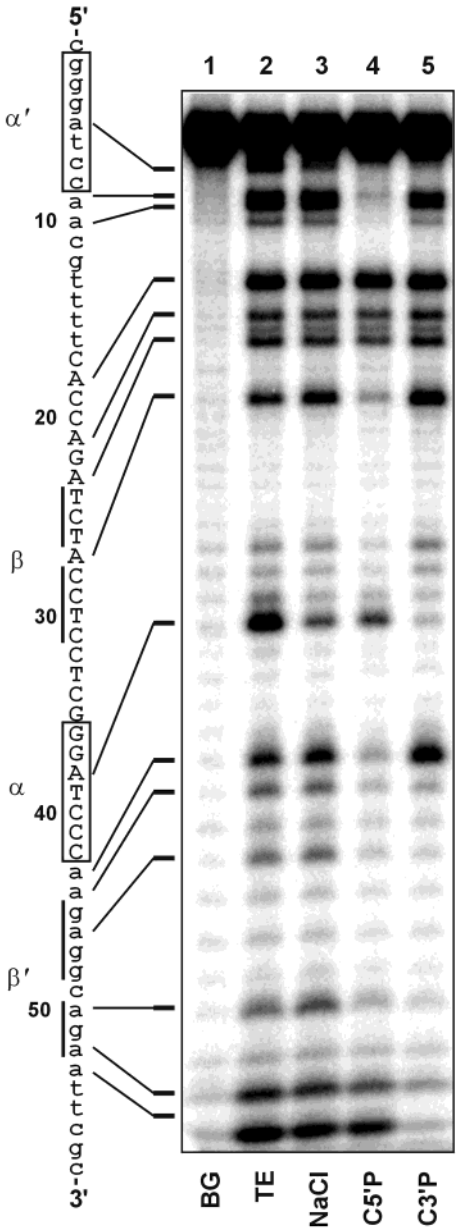

FIGURE 4: Cleavage patterns of U-4 nucleotides accessible to DEPC modification (modification protection experiment). 3′-^{32}P-labeled U-4 at 10 μM was prepared in TE buffer (lane 2) or in TE buffer with 200 mM NaCl (lanes 3−5). The U-4 samples in lanes 4 and 5 were respectively perturbed by the additions of 15 μM C5′P and 15 μM C3′P. Samples for lanes 2−5 were heated and cooled before reaction with DEPC. The modified DNA was cleaved by piperidine and separated on 12% sequencing gels. Lane 1 shows the background cleavage of U-4 that was heated and cooled in TE buffer but without DEPC treatment.

C5′P. As in the case of U-4, the U-4:C5′P hybrid showed concentration-dependent mobility on the gel (Figure 3A, lanes 9−12), indicating that this hybrid retained the ability to form a multimer. However, the spread of mobilities in Figure 3A, lanes 10 and 11, as compared with Figure 3A, lanes 2 and 3, probably indicated that the interchange between structures was significantly slower at moderate U-4 concentrations (1−0.1 μM) when the 5′-end was blocked. The U-4:C5′P hybrid formed discrete bands on the gel only at extreme DNA concentrations; a dimer (or multimer) was favored at high concentrations (Figure 3A, lane 9), and a monomer was dominant at low concentrations (Figure 3A,

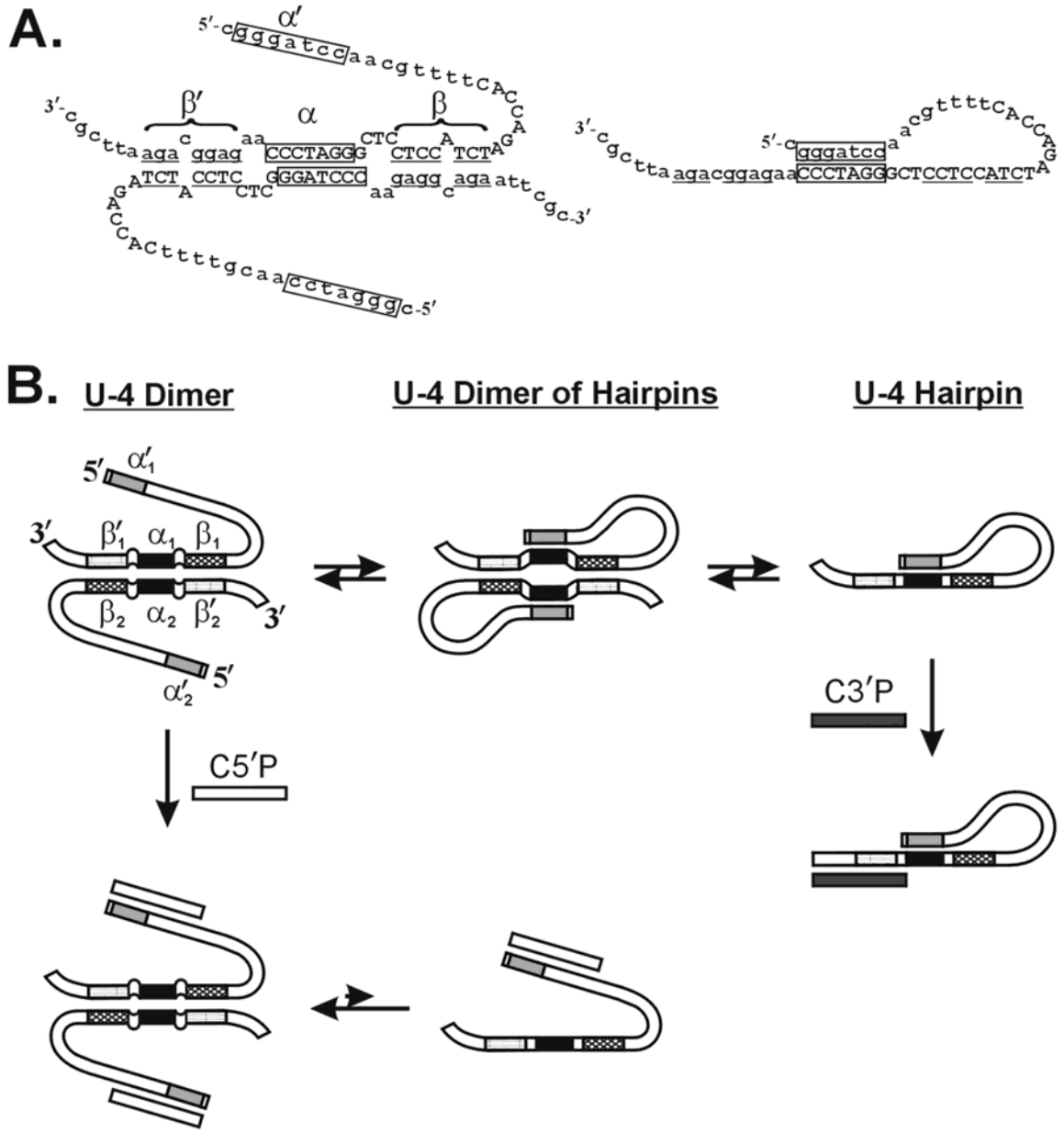

FIGURE 5: Schematic drawings of U-4 structures. (A) Proposed structures of U-4 dimer (left) and monomer hairpin (right). (B) Illustration of inferred U-4 structures and the effects of oligomer primers (C5′P and C3′P) on the structural interchange. See Table 1 for definitions of the segments α, α′, *β*, and *β*′; different subscripts are used to distinguish the two strands.

lane 12). In melting profiles, the first transition of the U-4:C5′P hybrid was further raised to 42.2 °C (Table 3), showing that the interstrand interactions were strengthened [i.e., the hybrid was more restrained to a dimer (or multimer) state]. Therefore, blocking the 5′-end promoted formation of a higher order structure.

A major transition at 59−60 °C for the U-4:C3′P and U-4:C5′P hybrids, and also for the U-4t42:C5′P hybrid (Table 3), was consistent with the melting of the primers from the ends of U-4 or U-4t42.

A similar set of structural transitions involving two states of a DNA dodecamer has been reported (*29*). Successive low- and high-temperature melting transitions were observed, corresponding to conversions from mismatched duplex (dimer) to hairpin and then to random coil, respectively. Interestingly, the high-temperature transition of the dodecamer was not detected by UV (but could be observed by circular dichroism), different from the case of U-4.

To summarize, the two constant ends of U-4 play opposite roles in the structural equilibrium: the 3′-end is required for the formation of a multimeric structure, while an available 5′-end facilitates the transition to a monomeric state.

DEPC Modification Protection of U-4 Structures. The U-4 structures were further investigated by DEPC chemical modification (*30*). DEPC reacts to carbethoxylate the N^7 position of adenine, and to some extent that of guanine, when the bases are not stacked within a duplex (*30*−*32*). As shown in Figure 4, the U-4 structure(s) formed in 200 mM NaCl provided only a limited protection in the α and α′ segments, mostly at G37 and A38 (compare lane 3 with lane 2 in Figure 4), presumably because most nucleotides became exposed during the transient interchange of structures. To restrict base pairing to discrete structures, we took advantage of the effect of primer hybridization. When hybridized with C5′P, which blocked the 5′-end, U-4 was restrained to a higher order structure (see lane 9 of Figure 3A). As shown in lane 4 of

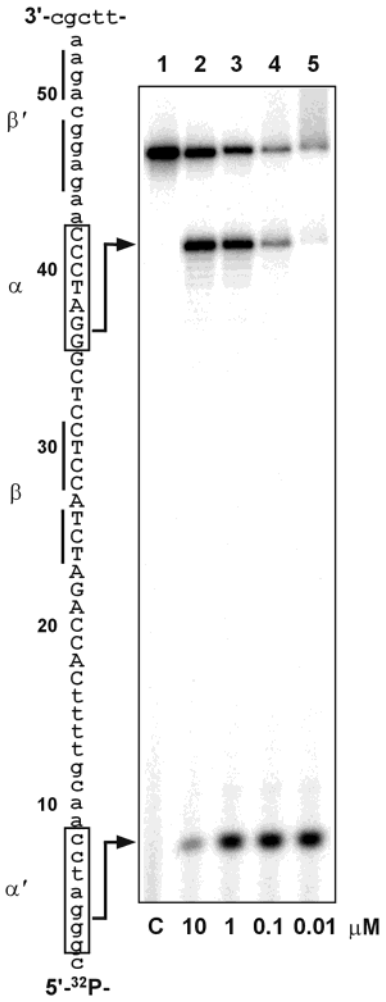

FIGURE 6: *BamH* I cleavage of U-4 structures. 10-fold serial dilutions of U-4 were prepared in 150 mM NaCl with a fixed amount of 5′-^{32}P-labeled U-4 in each dilution. Samples were heated and cooled before *BamH* I restriction enzyme digestion. Cleavage products were resolved in 20% sequencing gels. The U-4 concentrations are shown at the bottom of the figure. Lane 1 is a control sample of U-4 without enzyme digestion.

Figure 4, the entire 5′-end constant region was protected from modification, as expected. In addition, A27 (in β), G35-A38 (mostly in α), and A43-A50 (mostly in β') were also protected. These results were best explained by the formation of a dimer, as modeled in Figure 5A (left panel), plus the binding of a C5′P oligomer. On the other hand, when U-4 was hybridized with C3′P, which blocked the 3′-end and resulted in a monomeric form of U-4 (see lanes 5−8 of Figure 3A), the results showed that only A5 (in α') and G37 and A38 (in α) were protected, in addition to the entire 3′-end constant region (lane 5 of Figure 4). These results were satisfied by a hairpin model, as shown in Figure 5A (right panel), plus the binding of a C3′P oligomer. Such a hairpin structure was also predicted by *mfold*, a web-based folding program (http:// www.bioinfo.rpi.edu/applications/mfold/old/dna/; ref *33*). The bases in the internal loops of the U-4 dimer or in the large loop of the hairpin (Figure 5A) may interact through noncanonical base pairing as summarized by Chou et al. (*34*), but we have not considered such substructure details in our measurements.

The DEPC modification protection data did not support the existence of an alternative hairpin formed via base pairing of the β and β' segments in 200 mM NaCl (Figure 4, lane 3). This alternative hairpin structure would have the same number (seven) of base pairs in the stem region as the proposed hairpin but would be less stable due to the discontinuous nature and reduced G/C content of the stem.

Cleavage of U-4 Structures by BamH I. The palindromic nature of the α segment (GGATCC, see Table 1) meant that a *BamH* I restriction site was potentially formed in the proposed models of both dimer and hairpin (see Figure 5A). Therefore, existence of the structures could be further verified and distinguished by *BamH* I restriction cleavage of U-4 ^{32}P-labeled at the 5′-end. As shown in Figure 6, a large ^{32}P-labeled fragment prevailed at 10 μM (Figure 6, lane 2), corresponding to cleavage at the α segment, whereas a short fragment began to dominate at lower U-4 concentrations (Figure 6, lanes 3−5), corresponding to cleavage at α'. The results were consistent with our models. U-4 tends to form a dimer at high concentrations, and *BamH* I mainly cuts at α on both strands. A hairpin is the major structure at low concentrations, and both α and α' are simultaneously cleaved to generate a short-labeled fragment. Since the sequence is labeled at the 5′-end, only the shorter fragment is visible with the latter cleavage.

Model of the Dynamic Interchange of U-4 Structures. As shown previously, U-4 did not exist in an exclusive dimer or monomer state under most conditions. Therefore, we propose a dynamic model of the structural changes. In this model, illustrated in Figure 5B, a symmetrical dimer is formed through discontinuous base pairs (see Figure 5A for detailed structures). The α'_1 segment from strand 1 invades the paired region to form a hairpin and displaces α_2 of strand 2, which then pairs to α'_2 on the same strand resulting in a transient dimer of hairpins. Owing to the reduced number of base pairs, the dimer of hairpins is easily dissociated to monomers. This process is reversible, but the equilibrium will be shifted by oligonucleotide hybridization at the 3′- or 5′-ends. When hybridized with C3′P, which occupies part of the dimerization region (β'), U-4 is trapped in the hairpin state. On the other hand, blocking the 5′-end by C5′P inhibits the invasion of α' and thus stabilizes the dimer.

End Boundaries of U-4 for g5p Binding. To assess the importance of the 3′- and 5′-ends in binding g5p, we determined end boundaries of the portion of the U-4 sequence that retained high-affinity binding for g5p. ^{32}P-labeled U-4 was partially cleaved to generate variable lengths of fragments that were all labeled at the same end; this library of fragments was then titrated with g5p. The bound DNA fragments in the intermediate and saturated complexes were recovered from agarose gels and analyzed. As shown in Figure 7A, the boundary at the 3′-end was clearly positioned at A43 or C42 (Figure 7, lanes 3 and 4); C42 was required only at the lowest g5p concentration (P/N = 0.03; Figure 7, lane 3). In contrast, fragments of all lengths were bound and saturated at high concentrations of g5p (Figure 7, lanes 5 and 6).

For the 5′-end boundary, only a deletion of the first two nucleotides (5′-CG-) could be tolerated in the formation of the intermediate complex (Figure 7B, lanes 3 and 4). There was no obvious 5′-end boundary in fragments that were isolated from the saturated complexes (Figure 7B, lanes 5 and 6). Therefore, the high-affinity site of U-4 was within

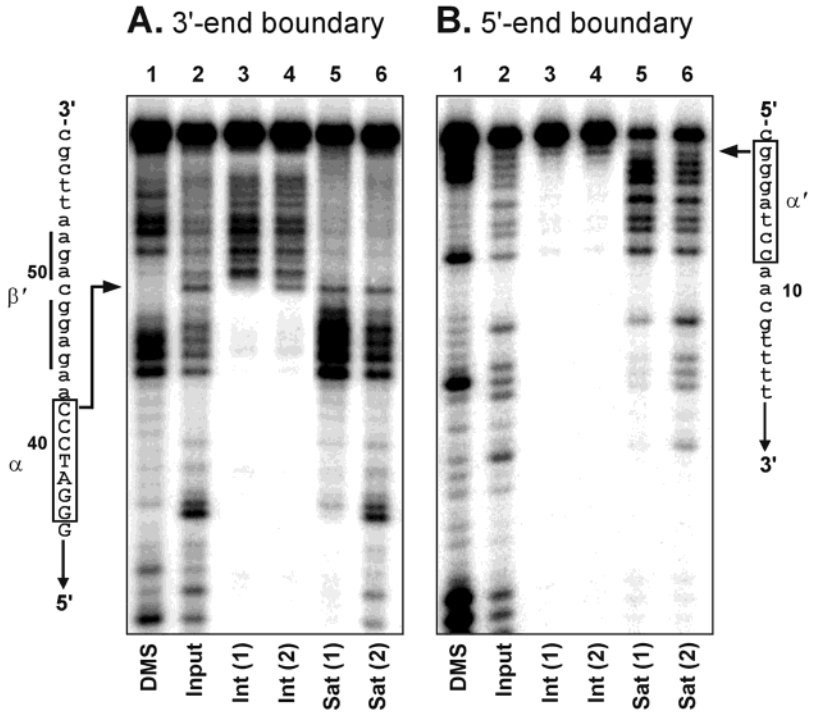

FIGURE 7: End boundaries of the U-4 sequence in the intermediate complex with g5p. U-4 fragments were ^{32}P-labeled at the 5′- and 3′-ends to respectively determine the 3′-end (A) and 5′-end (B) boundaries of the sequence required to form the intermediate complex with g5p. U-4 fragments were prepared (see Experimental Procedures for details) in 200 mM NaCl and titrated with g5p at 37 °C. Samples were separated on agarose gels. The intermediate and saturated complexes were isolated from the gel, and the DNA was extracted and resolved on 12% (A) or 8% (B) sequencing gels. For both panels: lane 1, DMS-cleaved U-4 as markers; lane 2, the mixture of U-4 fragments without g5p; lanes 3 and 4, intermediate complexes at P/N = 0.03 and 0.06, respectively; lanes 5 and 6, saturated complexes at P/N = 0.12 and 0.18, respectively. The respective boundaries are indicated at the sides. Data are shown for fragment concentrations of ~3 μM. Results were the same for fragment concentrations equivalent to ~3 or ~17 μM of the full-length U-4 sequence.

the sequence spanning G3-C41 and essentially ended at α′ and α. The results implied that a hairpin structure closed by the α:α′ stem was a minimal requirement for high-affinity binding and for the formation of the intermediate complex with g5p. This hypothesis was tested by the modification interference experiments below.

DEPC and DMS Modification of U-4 and Interference of g5p Binding. U-4 was partially modified by DEPC or DMS treatment, and the sequences that still formed intermediate or saturated complexes with g5p were isolated and assayed as described in Experimental Procedures. As shown in Figure 8, two regions of the U-4 sequence isolated from the intermediate complex were completely absent of DEPC modification: G37-A38 in the α segment and G4-A5 in the α′ segment (compare lanes 3 and 4 with lane 2 in Figure 8; fragments cleaved in C1-G3 were not resolved on the gel), both in the expected base-paired region of a hairpin. DEPC modification did not prevent the g5p from forming saturated complexes with U-4, independent of the ability to fold into a hairpin structure (Figure 8, lane 5). DEPC introduces a bulky carbethoxyl group at the N^7 position of adenines and guanines (*31*) that are not involved in Watson−Crick base pairing. Thus, the interference effect was probably due to hindrance of the added bulky group on base stacking and pairing (especially for the short 7-bp duplex stem of U-4), resulting in reduced formation of the hairpin structure and reduction of binding affinity with g5p. Another possibility was that the added group disrupted the contact surface between g5p and DNA; that is, the N^7 sites of G37, A38, G4, and A5 might be required for binding of g5p to form the intermediate complex. To test this possibility, DMS modification interference was performed as for DEPC. DMS adds a smaller methyl group at the N^7 position of guanines that does not interfere with Watson−Crick pairing (*26*). The results showed that DMS methylation, even methylation of guanines in the α and α′ segments, did not interfere with formation of the g5p·U-4 intermediate complex (lanes 7 and 8 as compared with lane 6, Figure 8), indicating that the availability of the N^7 sites was not required for g5p binding. Moreover, it is not likely that the bulky DEPC group just sterically hindered the binding since the formation of a saturated g5p·U-4 complex was not affected by any DEPC modification (Figure 8, lane 5), and modifications in the proposed loop region (A9-G35) did not block formation of intermediate complexes (lanes 3 and 4, Figure 8). Therefore, these data supported the hypothesis that formation of a U-4 hairpin structure was a prerequisite for high affinity g5p binding.

Competitive Binding for g5p. To further demonstrate the importance of a hairpin structure on g5p binding, we compared the relative protein binding affinity of U-4t42 and its mutant counterparts (see Figure 9B for illustrations). U-4t42 was a minimal U-4 sequence motif that embraced the tentative hairpin structure. The sequence hp42 had shuffled base pairs in the stem region of the U-4t42 hairpin such that a hairpin structure was maintained (T_m = 55 °C at 200 mM NaCl). In contrast, ss42 was mutated in the stem region such that stable hairpins did not form (T_m < 15 °C at 200 mM NaCl). Competitive experiments were performed by mixing cold U-4t42, hp42, or ss42 with labeled U-4t42 prior to g5p binding. The EMSA results in Figure 9A showed that U-4t42 and hp42 had comparable g5p binding affinity

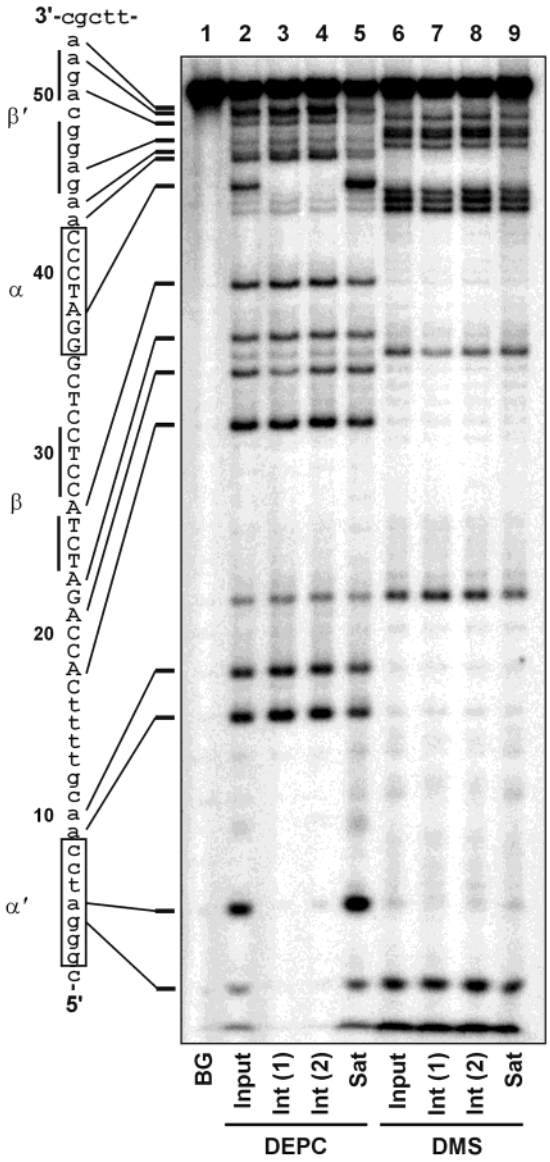

FIGURE 8: DEPC and DMS cleavage patterns of U-4 nucleotides required for formation of the g5p·U-4 intermediate complex (DEPC and DMS modification interference experiments). 5′-^{32}P-labeled U-4 was partially modified by DEPC (lanes 2−5) or DMS (lanes 6−9). The modified DNA was allowed to equilibrate into folded structures in 200 mM NaCl and was then titrated with g5p, all at 37 °C. Samples were separated on agarose gels. The intermediate and saturated complexes were isolated from the gels. The DNA was extracted, cleaved by piperidine, and resolved on 20% sequencing gels. Lane 1, background cleavage; lanes 2 and 6, modified U-4 without g5p; lanes 3 and 7, intermediate complexes at P/N = 0.06; lanes 4 and 8, intermediate complexes at P/N = 0.2; lanes 5 and 9, saturated complexes at P/N = 0.2.

(compare lanes 7−10 with 3−6, Figure 9A), whereas ss42 was much less competitive (compare lanes 11−14 with 3−6, Figure 9A). On the basis of these and similar competition experiments, the g5p-binding affinity of U-4t42 was estimated to be about 40-fold higher than that of ss42 (data not shown). These results demonstrated the important role of a hairpin structure in determining the high affinity g5p binding to U-4.

DISCUSSION

On the basis of their competitive ability to bind the g5p under physiologically relevant conditions (200 mM NaCl, pH 7.4, and 37 °C), DNA sequences that could form hairpin structures were SELEX-selected. Even though the g5p has been well-characterized as a simple ssDNA binding protein (*6*, *35*), we can now include DNA hairpins and G-quadruplexes (*21*, *22*) as being among the protein's preferred binding sites.

Is Dimerization of the Hairpins Important for g5p Binding? An interesting finding from the study of a representative selected sequence, U-4, was that it existed in interchangeable states of monomers and dimers. Moreover, the monomer formed a hairpin that was predominant when the 3′-end was blocked, while blockage of the 5′-end shifted the equilibrium toward the dimer by inhibiting the formation of a hairpin structure (through the α and α′ segments; see Figures 3 and 5). The affinity of g5p for U-4 was similar whether its 3′-end was intact, blocked (by a complementary oligomer), or deleted, whereas the affinity was significantly reduced when the 5′-end was blocked (data not shown). Therefore, the formation of a hairpin structure (in the form of monomers or dimers) appeared to be required for binding g5p with high affinity. The requirement for hairpin formation was supported by the DEPC modification interference and competitive binding experiments (see Figures 8 and 9, respectively). The sequence boundary experiments (see Figure 7) also suggested that a monomer of hairpins was sufficient for formation of an intermediate complex.

Since the results in Figure 7 were independent of U-4 concentration (see the figure legend), and showed no evidence for an effect of U-4 dimer formation on g5p binding, the g5p binding affinities for a hairpin monomer and a dimer under these conditions apparently were very similar. However, even a minor difference might be amplified during the rounds of competition-based selection. For example, eight out of 26 family I sequences from the eighth round of SELEX had an α segment (GGATCCC) that could form hairpin structures with 7-bp stems, while another nine family I sequences had α segments (GATCCC) that were one nucleotide shorter and presumably formed less stable 6-bp stems (see Table 1). In contrast, the corresponding numbers of analogous sequences from the sixth round of SELEX were one and 11, from among 22 sequences. (For the sixth round, 37 clones were sequenced, and 22 of these had motifs similar to those of the family I sequences listed in Table 1; data not shown.) This evolution between the sixth and the eighth round of SELEX indicated that g5p binding differentiated between hairpin structures having six and seven bp. That is, a small but significant difference can be amplified under iterative selection pressure. Similarly, the dimerization of hairpins might influence g5p binding and selection in subtle ways we have yet to appreciate.

Furthermore, intermolecular interactions that lead to preferential binding might not converge to obvious consensus sequences, and individual sequences in family I for which the putative β:β' pairing was less stable may not be able to form homodimers as did U-4. [e.g., The β:β' region of V-1 consisted of four complementary bases interrupted by two mismatches (see Table 1), and in fact, the β:β' sequence population was quite diverse throughout the whole SELEX procedure.] That is, at the high DNA concentrations (10−100 μM) used during the SELEX selection, intermolecular interactions between DNA molecules undoubtedly involved two or more different sequences, with the pyrimidine-rich

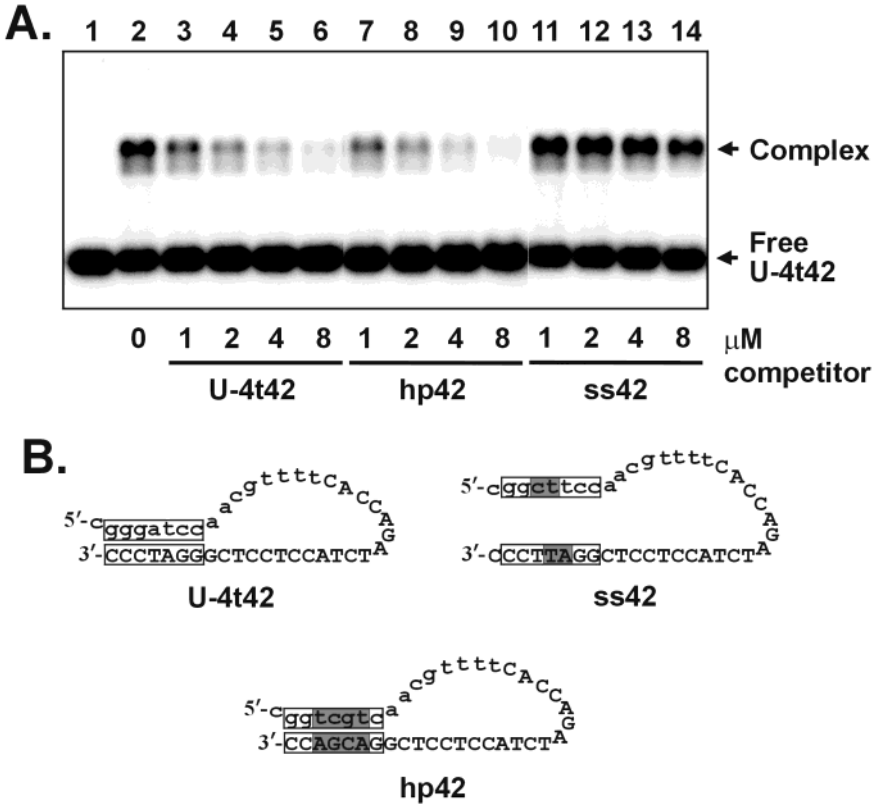

FIGURE 9: (A) EMSA of competitive binding for g5p. The 1 μM ^{32}P-labeled U-4t42 was mixed with buffer (lane 2) or with unlabeled U-4t42 (lanes 3−6), hp42 (lanes 7−10), and ss42 (lanes 11−14) as competitors. Concentrations of the competitors are shown at the bottom of the figure. The g5p was added to each sample to a final concentration of 6 μM. The mixtures were incubated at 37 °C for 15 min and subjected to electrophoresis on 2.5% agarose gels. Lane 1, ^{32}P-labeled U-4t42 only. (B) Schematic drawings of U-4t42, hp42, and ss42. The mutated nucleotides from U-4t42 are shaded.

region of the β segment of one sequence pairing with the β' segment of the purine-rich 3′ primer region of another sequence to form a dimer like the one shown in the left panel of Figure 5A. We conclude that the absence of a consensus β:β' sequence does not argue against the possible significance of dimer formation during the selection.

Another aspect makes the dimer model even more attractive. As shown in Figure 5B, the putative transient structure of a dimer of hairpins brings together four complementary and antiparallel strands with constrained, looped-out nucleotides. The strand orientation, bonding, and presence of constrained nucleotides of such a structure are reminiscent of the intrastrand G-quadruplex fold of I-3, which was found to be an initiation site for g5p binding (*22*). That is, in terms of g5p binding, there are convergent features of the G-quadruplex and the dimer of hairpins that rationalize how two distinctive DNA structures were derived from the two SELEX protocols.

Assembly of g5p on the U-4 Hairpin versus the I-3 Quadruplex. In one regard, g5p assembled on both the U-4 and the I-3 sequences in a sequential manner, leading to the formation of a saturated complex (at high P/N ratios) via an intermediate, unsaturated complex (at low P/N ratios; see Figure 2). However, the transition from the intermediate to saturated complex for I-3 occurred at relatively lower P/N ratios (P/N $\geq$ 0.07) than did the transition for U-4 (P/N $>$ 0.19). In the case of I-3, we previously showed that g5p first binds to the G-quadruplex core and subsequently saturates the antiparallel-orientated tails with only minimal alteration of the core structure (*22*). That is, the first three g5p dimers

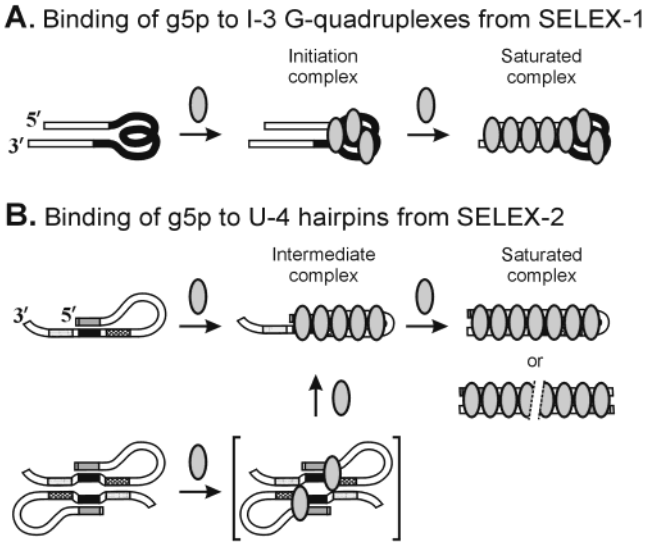

FIGURE 10: Schematic models for g5p binding to I-3 G-quadruplexes from SELEX-1 (A) and to U-4 hairpins from SELEX-2 (B). Ovals stand for g5p dimers. The variable region of I-3 including the G-quadruplex core is shown in black lines, and the constant tails are in open lines. Presentation of the U-4 structures is the same as in Figure 5. The duplex stem of the U-4 hairpin may or may not be bound by g5p in the intermediate complex.

bind nucleotides constrained within the G-quadruplex structure (*21*), and formation of this intermediate complex initiates cooperative saturation of the tails. This assembly process is illustrated in Figure 10A. In the case of U-4, we do not know as precisely the localization of the five g5p dimers that are involved in the intermediate complex. However, given the

requirement of a hairpin structure to form an intermediate complex, and the fact that the g5p has dyadic DNA-binding sites (*8*), the g5p is likely to bind to and saturate the restrained antiparallel strands in the loop region of the hairpin, leaving mainly the 3′-primer tail unbound. This assembly (the intermediate complex) is apparently a closed complex, in the sense that it is stable over a wide range of P/N ratios and does not facilitate or initiate the binding of further g5p dimers. The closed complex can be opened to form a saturated complex only when the g5p concentration is raised to a critical point (e.g., P/N > 0.19, see Figure 2A), probably to melt the duplex stem and saturate all of the nucleotides, either by rearranging the hairpin to align the tails or by bringing two U-4 antiparallel strands together to form the complex. The binding of g5p to a U-4 hairpin is illustrated in Figure 10B (top line). Assuming that the binding is facilitated by formation of a dimer of hairpins, we speculate that the dimerization might result in a higher negative charge density and a more favorable presentation of constrained nucleotides for g5p binding (Figure 10B, bottom line).

This model, together with the strong cooperative binding of g5p, suggests that hairpins with large loops that will accommodate a maximum number of g5p dimers should outcompete those with smaller loops during the SELEX-2 selection of intermediate complexes. In agreement with this suggestion, the sequences of family I (see Table 1) did have a maximal loop size of the hairpins, in that the α segment was located at or almost at the boundary between the variable and the 3′-primer regions, and the α′ segment was only zero to two nucleotides away from the very 5′-end of the sequences. Why were hairpins with equivalently large loops not selected that had free 5′-constant region tails and nucleotides at the 5′-end of the variable region that paired with nucleotides in the 3′-constant region (e.g., GAATTCG at positions 51−57)? The reduced stability of a hairpin with a majority of A:T pairs is a likely answer. Another possibility is that there is some sequence specificity of binding of the g5p to the α:α′ dsDNA stem. This was less likely because the competition experiments showed comparable binding affinities of g5p to U-4t42 and hp42, the latter of which has scrambled base pairs in the stem region of the hairpin (see Figure 9). A similar result was reported from another SELEX study with C5 protein of RNase P, in which a sequence change in the stem region of a SELEX-selected RNA hairpin did not affect the target protein binding (*36*).

Biological Relevance. From the study of interactions between the g5p and the SELEX-selected sequences, we have found that two distinct types of DNA structures, G-quadruplexes and hairpins, can provide high-affinity sites for g5p binding. Of these two, hairpins may be the more relevant in the biological functioning of the g5p. A search of the fd genome yielded only six G_{4-5} blocks, two of which are located in or near the gene 2 operator responsible for g5p-mediated translational repression (*14*). It seems improbable that these G_4 blocks could participate in a G-quadruplex to affect translation because at least two mRNA strands would have to be involved. In contrast, based on *mfold* calculations, there are more than 100 hairpins distributed throughout the fd DNA genome, although only ∼10% of them have loop sizes that are at least as large as that of U-4. Moreover, *mfold* calculations with the RNA yield extensive stem−loop structures, with mismatches and bulge loops, that involve the mRNA leader sequences for all five viral genes 1, 2, 3, 5, and 10 that are controlled by g5p. Favorably constrained nucleotides in such complex RNA stem−loop structures may provide excellent targets for g5p binding, although this remains to be tested. The involvement of hairpins in other biological functions of the g5p, such as providing immunity to Ff superinfection, is also possible. In any case, the finding that g5p assembles on hairpin-constrained nucleotides points to a new dimension of how this small ssDNA-binding protein might achieve its sequence-specific binding functions.

REFERENCES

1. Beck, E., and Zink, B. (1981) *Gene 16*, 35−58.
2. Rapoza, M. P., and Webster, R. E. (1995) *J. Mol. Biol. 248*, 627−638.
3. Schaller, H. (1979) *Cold Spring Harbor Symp. Quant. Biol. 43 Pt 1*, 401−408.
4. Zinder, N. D., and Horiuchi, K. (1985) *Microbiol. Rev. 49*, 101−106.
5. Gray, C. W. (1989) *J. Mol. Biol. 208*, 57−64.
6. Model, P., and Russel, M. (1988) in *The Bacteriophages* (Calendar, R., Ed.) pp 375−390, Plenum Press, New York.
7. Olah, G. A., Gray, D. M., Gray, C. W., Kergil, D. L., Sosnick, T. R., Mark, B. L., Vaughan, M. R., and Trewhella, J. (1995) *J. Mol. Biol. 249*, 576−594.
8. 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.
9. Terwilliger, T. C. (1996) *Biochemistry 35*, 16652−16664.
10. Kansy, J. W., Clack, B. A., and Gray, D. M. (1986) *J. Biomol. Struct. Dyn. 3*, 1079−1110.
11. 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.
12. Bulsink, H., Harmsen, B. J., and Hilbers, C. W. (1985) *J. Biomol. Struct. Dyn. 3*, 227−247.
13. Mou, T.-C., Gray, C. W., and Gray, D. M. (1999) *Biophys. J. 76*, 1537−1551.
14. Michel, B., and Zinder, N. D. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 4002−4006.
15. Zaman, G., Smetsers, A., Kaan, A., Schoenmakers, J., and Konings, R. (1991) *Biochim. Biophys. Acta 1089*, 183−192.
16. Michel, B., and Zinder, N. D. (1989) *Nucleic Acids Res. 17*, 7333−7344.
17. Oliver, A. W., and Kneale, G. G. (1999) *Biochem. J. 339*, 525−531.
18. Oliver, A. W., Bogdarina, I., Schroeder, E., Taylor, I. A., and Kneale, G. G. (2000) *J. Mol. Biol. 301*, 575−584.
19. Ellington, A. D., and Szostak, J. W. (1990) *Nature 346*, 818−822.
20. Tuerk, C., and Gold, L. (1990) *Science 249*, 505−510.
21. Wen, J.-D., Gray, C. W., and Gray, D. M. (2001) *Biochemistry 40*, 9300−9310.
22. Wen, J.-D., and Gray, D. M. (2002) *Biochemistry 41*, 11438−11448.
23. Gray, D. M., Hung, S. H., and Johnson, K. H. (1995) *Methods Enzymol. 246*, 19−34.
24. Laemmli, U. K. (1970) *Nature 227*, 680−685.
25. Steinberg, T. H., Jones, L. J., Haugland, R. P., and Singer, V. L. (1996) *Anal. Biochem. 239*, 223−237.
26. Maxam, A. M., and Gilbert, W. (1980) *Methods Enzymol. 65*, 499−560.
27. Schneider, D. J., Feigon, J., Hostomsky, Z., and Gold, L. (1995) *Biochemistry 34*, 9599−9610.
28. Gold, L., Polisky, B., Uhlenbeck, O., and Yarus, M. (1995) *Annu. Rev. Biochem. 64*, 763−797.
29. Davis, T. M., McFail-Isom, L., Keane, E., and Williams, L. D. (1998) *Biochemistry 37*, 6975−6978.

30. Merryman, C., and Noller, H. F. (1998) in *RNA:Protein Interactions* (Smith, C. W. J., Ed.) pp 237−253, Oxford University Press, New York.
31. Peattie, D. A. (1979) *Proc. Natl. Acad. Sci. U.S.A. 76*, 1760−1764.
32. Weeks, K. M., and Crothers, D. M. (1993) *Science 261*, 1574−1577.
33. SantaLucia, J., Jr. (1998) *Proc. Natl. Acad. Sci. U.S.A. 95*, 1460−1465.
34. Chou, S. H., Chin, K. H., and Wang, A. H. (2003) *Nucleic Acids Res. 31*, 2461−2474.
35. Alberts, B., Frey, L., and Delius, H. (1972) *J. Mol. Biol. 68*, 139−152.
36. Lee, J. H., Kim, H., Ko, J., and Lee, Y. (2002) *Nucleic Acids Res. 30*, 5360−5368.

BI030177G