SELEX Selection of High-Affinity Oligonucleotides for Bacteriophage Ff Gene 5 Protein†

Jin-Der Wen, Carla W. Gray, and Donald M. Gray*

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

Received January 17, 2001; Revised Manuscript Received May 3, 2001

ABSTRACT: The Ff gene 5 protein (g5p) is a cooperative ssDNA-binding protein. SELEX was used to identify DNA sequences favorable for g5p binding at physiological ionic strength (200 mM NaCl) and 37 °C. Sequences were selected from a library of 58-mers that contained a central variable segment of 26 nucleotides. DNA sequences selected after eight rounds of SELEX were mostly G-rich, with multiple copies of CPuGGPy, TPuGGGPy, and/or PyPuPuGGGPy motifs. This was unexpected, since g5p has higher binding affinities for polypyrimidine than for polypurine sequences. The most recurrent G-rich sequence, named I-3, was found to have g5p-binding properties that were correlated with a structural transition. At 10 mM NaCl, I-3 existed in a single-stranded form that was saturated by g5p in an all-or-none fashion. At 200 mM NaCl, I-3 existed in a structured form that showed CD spectral features of G-quadruplexes. The g5p binding affinity for this structured form of I-3 was >100-fold higher than for the single-stranded form. Moreover, the structured I-3 was saturated by g5p in two steps, the first of which was the formation of an apparent initiation complex consisting of one I-3 strand and about three g5p dimers. Nuclease S1 footprinting and other experiments showed that g5p molecules in the initiation complex at 200 mM NaCl were bound directly to the G-rich variable segment and that the structure of I-3 was retained after saturation by g5p. Thus, G-rich motifs may form structures favorable for initiation of g5p binding and also provide the actual g5p-binding sites.

The genome of the Ff phages[1] comprises 11 tightly packed genes and an intergenic region between genes 4 and 2 (*1*). The Ff g5p is a single-stranded DNA binding protein of monomer MW 9690 (87 amino acid residues) that exists as a stable dimer even at a concentration as low as 1 nM (*2*). The g5p dimer has 2-fold rotational symmetry and cooperatively saturates the Ff ssDNA genome by binding antiparallel-stranded nucleotides in its dyadic DNA-binding sites (*3*−*5*). There are three modes of binding, in which 4, 3, or 2.5 nucleotides are bound per g5p monomer (i.e., $n = 4$, 3, or 2.5; *6*, *7*). The $n = 4$ mode of binding is the dominant mode when g5p is present at P/N ratios ≤ 0.25. However, there is virtually no information available on how g5p initiates cooperative binding to the viral genome under physiological conditions. The ssDNA complexed with g5p does not serve as a template for dsDNA synthesis and is a precursor for virion assembly (*1*).

Owing to its biological functions in saturating the viral genome, g5p has been most studied for its non-sequence-specific binding properties (*3*, *8*). However, g5p is known to have marked differences in binding affinities (Kω) for synthetic single-stranded polynucleotides (*9*, *10*). A polynucleotide that is more stacked, like the polypurine poly[d(A)], binds with lower affinity than do the less stacked polypyrimidines poly[d(T)] and poly[d(C)]. Moreover, g5p has structure- and sequence-specific binding functions of biological importance. (a) The complexes isolated from cells frequently have three or four branches, suggestive of local preferential initiation at more than one site (C. W. Gray, unpublished results). (b) A (G+C)-rich hairpin of 32 base pairs in the intergenic region is oriented at one end of the g5p·ssDNA intracellular complex (*11*) and maintains the same orientation in the mature virus after being packaged (*12*). (c) The g5p inhibits the "−" strand synthesis and synthesis of replicative form (RF) dsDNA, but not merely by sequestering genomic "+" strands. Fulford and Model (*13*) proposed a competitive melting by g5p and stabilization by g2p of hairpins at the "−" strand origin as a switching mechanism that controls synthesis of RF dsDNA. (d) Fulford and Model (*13*) also showed that low levels of g5p provide immunity to Ff superinfection, inconsistent with simple saturation of the infecting ssDNA by g5p. (e) The g5p binds

† 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. Department of Molecular and Cell Biology, Mail Stop FO 3.1, The University of Texas at Dallas, Box 830688, Richardson, TX 75083-0688; (972) 883-2513; FAX (972) 883-2409; e-mail: dongray@utdallas.edu.

[1] Abbreviations: aptamer, a nucleic acid sequence selected to have high affinity for a protein or other substance; CD, circular dichroism; 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; Kω, the intrinsic binding constant (K) times a cooperativity factor (ω); PCR, polymerase chain reaction; P/N, the [protein monomer]/[nucleotide] molar ratio; SDS−PAGE, sodium dodecyl sulfate−polyacrylamide gel electrophoresis; SELEX, Systematic Evolution of Ligands by Exponential enrichment; ssDNA, single-stranded DNA; TAE buffer, 40 mM Tris-acetate, pH 8.3, 1 mM EDTA; TBE buffer, 90 mM Tris-borate, pH 8.3, 2 mM EDTA; TE buffer, 10 mM Tris-HCl, pH 7.4, 1 mM EDTA.

10.1021/bi010109z CCC: $20.00
Published on Web 07/12/2001

to the mRNA leader sequences of genes 1, 2, 3, 5, and 10 to regulate their translation (*14*). Binding sites in the gene 2 and 3 leader sequences contain G-rich blocks of four or five purines surrounded by blocks of pyrimidines. (f) The g5p has been shown to repress the translation of gene 2 mRNA both in vitro and in vivo (*15*, *16*). Michel and Zinder (*17*) have defined a sequence of 16 nucleotides in the gene 2 mRNA that is required in vivo for repression by g5p. They also found that, in vitro, an RNA carrying this sequence is at least 10-fold higher in affinity for g5p binding than is an RNA lacking it. The preferential binding of g5p to an RNA carrying the 16-mer sequence is affected by mutations that abolish gene 2 translational repression in vivo (*18*). (g) A direct measurement by mass spectrometry shows that two g5p dimers bind to a DNA analogue of this 16-mer RNA, bending it to form a hairpin (*19*). Therefore, as shown by these examples, g5p plays a role in regulation of viral DNA synthesis and viral gene expression, probably through structure- and/or sequence-specific binding.

The SELEX methodology was originally developed to select from an in vitro library (a pool of RNA sequences with a randomized region and two flanking constant sequences) those sequences having high affinity for a protein (*20*) or immobilized dyes (*21*). SELEX has also been applied to the selection of DNA molecules from double- and single-stranded DNA libraries. The targets that have been used for SELEX include (a) nucleic acid-binding proteins, (b) proteins that are not thought to bind nucleic acids naturally, and (c) small molecules such as nucleotides, amino acids, cofactors, and dyes (*22*−*26*). The first example of SELEX using an ssDNA library was with thrombin, a protease that was considered not to interact physiologically with nucleic acids (*27*). Aptamers of ssDNA that bind thrombin with high affinities display a highly conserved region of 14−17 nucleotides, of which one typical sequence folds into a unimolecular quadruplex containing two G-quartets, as determined by NMR spectroscopy (*28*, *29*).

In this report, we present the first application of SELEX using a cooperative ssDNA binding protein, the Ff g5p. The g5p binds with a large, positive cooperativity factor, ω, so that the protein tends to saturate all nucleic acid sequences. Nevertheless, our results show that the SELEX strategy can be used to efficiently select high-affinity ssDNA sequences. The gist of our results is that the initiation of cooperative binding, when the binding is least dependent on ω, can be very dependent on the ssDNA sequence and structure.

EXPERIMENTAL PROCEDURES

SELEX, PCR, and Sequencing. The SELEX procedure applied in this paper was derived from that used by Gold and co-workers (*30*). A synthesized library of 58-mer ssDNA, called PV-58, was used for SELEX selection. PV-58 DNA was synthesized with the following sequence: 5′-CGG-GATCCAACGTTTT-N_{26}-AAGAGGCAGAATTCGC-3′ (Oligos Etc.). A, G, C, and T were randomly incorporated in the central 26 nucleotides (N_{26}). The two flanking constant 16-mer sequences were used as primer annealing sites for PCR amplification. The g5p was isolated and purified as in previous work (*7*, *10*).

For the initial selection, 1 nmol (6×10^{14} sequences) of PV-58 (a small portion was ^{32}P-labeled at the 5′ end with T4 polynucleotide kinase; see below) was incubated with g5p in 200 mM NaCl at 37 °C for 15 min in TE buffer (10 mM Tris-HCl, pH 7.4, 1 mM EDTA). The amount of g5p was adjusted to give a selection ratio (complexed DNA/total DNA) of 0.005 to 0.05 for each round of selection. G5p•DNA complexes were separated from free DNAs on the basis of EMSA (see below), except that the gels were not fixed. Instead, the resulting wet gel was exposed briefly to a storage phosphor screen (Molecular Dynamics). Superimposing the gel with the printed image enabled the position of complexed DNA to be located. The band corresponding to a saturated complex was excised from the wet agarose gel and the excised gel slice was liquefied by treatment with β-agarase (FMC BioProducts) at 45 °C for 1−2 h. The g5p-bound DNA was extracted by phenol/chloroform and then ethanol-precipitated.

PCR was performed to amplify the selected DNA with 5′ primer (5′-CGGGATCCAACGTTTT-3′) and biotin-conjugated 3′ primer (biotin-5′-GCGAATTCTGCCTCTT-3′) (Midland) under the following conditions: 95 °C for 2 min; 16 cycles at 95 °C for 1 min, 50 °C for 1 min, and 72 °C for 2 min; the final extension was at 72 °C for 10 min. [α^{32}P]-dCTP (ICN) was included in the PCR mixture to internally label the DNA. PCR products were purified in agarose gels with the *QIAEX II* kit (QIAGEN). To isolate the target ssDNA (extended with the 5′ primer), an immobilized streptavidin agarose bead matrix (Pierce) was added to bind the purified biotin-conjugated dsDNA, which was then denatured by 0.12 N NaOH at 37 °C for 15 min. Since the biotin−streptavidin interaction was not disrupted by this alkaline solution, only the target ssDNA was released from the matrix (*30*). The ssDNA was finally recovered by ethanol precipitation. Generally, 200−300 pmol of the enriched ssDNA was used for the following round of SELEX.

DNAs from the fourth, sixth, and eighth rounds of selection, as well as the original sample of PV-58, were cloned with the *TOPO TA Cloning* kit (Invitrogen) and sequenced with the *fmol DNA Sequencing System* (Promega).

EMSA. The I-3 DNA sequence (5′-CGGGATCCAACGTTTT-GGGGTCAGGCTGGGGTTGTGCAGGTC-AAGAGGCA-GAATTCGC-3′) (Oligos Etc.) was ^{32}P-labeled at the 5′ end using T4 polynucleotide kinase (Invitrogen) and [γ^{32}P]ATP (ICN). For titrations with g5p, 1 μM of labeled I-3 was incubated with 0−21 μM of g5p at 37 °C for 15 min in TE buffer containing 10 or 200 mM NaCl. Mixtures were loaded on 2.5% low-melting agarose gels (FMC BioProducts) in TAE buffer (40 mM Tris-acetate, pH 8.3, 1 mM EDTA), followed by electrophoresis at 8−10 V/cm for 80 min. DNA that was complexed with g5p had reduced electrophoretic mobility and was shifted to higher molecular-weight positions on gels. The gels were fixed with 10% acetic acid/50% methanol for 2 h, dried, and exposed to a storage phosphor screen. The bands were quantitated and analyzed with ImageQuant, v. 5.0 (Molecular Dynamics).

For competition experiments, 1 μM of ^{32}P-labeled I-3 was incubated in TE buffer with 14 μM of g5p in the absence or presence of 2 μM of an unlabeled competitor, I-7, at 37 °C for 15 min. I-7 (Oligos Etc.) differed from I-3 in that its central 26 nucleotide segment had the sequence 5′-GTGC-CACCCTCCTCTCTTGTTCTTGT-3′. The NaCl concentrations were 10, 50, 100, and 200 mM. After electrophoresis

of the samples, the gels were fixed and quantitated as described above.

To determine the apparent binding affinities, $K\omega_{app}$, for a g5p dimer, EMSA titrations of ssDNA (I-3 or PV-58) with g5p were quantitated and the apparent g5p binding affinity was determined as $K\omega_{app} = 1/(2L)$, where L is the free g5p dimer concentration at which 50% of labeled DNA was saturated with g5p. This binding affinity is for binding of a g5p dimer in one orientation, and the binding is assumed to be all-or-none (*7*, *10*). Free g5p concentrations were estimated using the assumption that four nucleotides were bound per g5p monomer in the $n = 4$ binding mode.

CD Measurements. CD spectra were measured in a Jasco model J710 spectropolarimeter, smoothed, and plotted at 1-nm intervals as molar CD ($\epsilon_L - \epsilon_R$) in units of M^{-1} cm^{-1}, per mole of nucleotide, as in previous work (*10*), except that spectra were smoothed by the method of fast Fourier transformation (Jasco).

Primer-Annealing and Serial Dilution Experiments. To determine the strand stoichiometry of I-3 in its free state and in the initiation complex with g5p, 1 μM of ^{32}P-labeled I-3 was incubated with 0−4 μM of the 3′ primer in 200 mM NaCl at 37 °C for 30 min. For the formation of complexes, g5p was added to give a final concentration of 7 μM for the final 15 min of the incubation. Control experiments were performed by reverse additions, preincubating I-3 with g5p for 15 min prior to the addition of the 3′ primer for another 15-min incubation. The 3′ primer sequence was the one described above but without conjugated biotin. Mixtures were subjected to electrophoresis (10 V/cm) in 12% polyacrylamide (acrylamide/bis = 19:1) gels in TBE buffer (90 mM Tris-borate, pH 8.3, 2 mM EDTA) for 4 h. Gels were fixed with 10% acetic acid for 10 min, dried, and exposed to a storage phosphor screen.

For serial dilutions, I-3 was diluted in 10 or 200 mM NaCl to final concentrations of 4, 1, 0.25, 0.063, and 0.016 μM strand. A trace amount of ^{32}P-labeled I-3 was added to each dilution. Samples were heated at 90 °C for 3 min, cooled at room temperature, and then subjected to electrophoresis as described above.

Stoichiometry of Protein and DNA in g5p•I-3 Complexes. To determine the stoichiometric ratio of protein to DNA, g5p and I-3 in the complexes were quantitated together by the following procedure: ^{32}P-labeled I-3 was mixed with g5p at 10 mM NaCl (6 μM I-3 per 32 μM g5p) or 200 mM NaCl (12.5 μM I-3 per 80 μM g5p). Mixtures were incubated at 37 °C for 15 min and electrophoresed in low-melting agarose gels as described above. The bands of g5p·I-3 complexes were isolated and heated with SDS loading buffer at final concentrations of 50 mM Tris-HCl, pH 6.8, 100 mM dithiothreitol, 2% SDS, 0.1% bromophenol blue, and 10% glycerol. The melted agarose solution was loaded on a 15% polyacrylamide gel, and SDS−PAGE was performed by the method of Laemmli (*31*). A series of standard amounts of g5p and ^{32}P-labeled I-3 were also run in the same gel. For g5p quantitation, the gel was stained with SYPRO Red (BioWhittaker Molecular Applications; *32*) for 2 h, destained with 7.5% acetic acid for 10 min, and scanned with a STORM 860 (red fluorescence mode; Molecular Dynamics). For I-3 quantitation, the same gel was dried and exposed to a storage phosphor screen. Calibration curves were plotted with the standards, and the respective amounts of g5p and I-3 in the complexes were calculated.

Quantitative Nuclease S1 Footprinting. To determine the initiation sites on structured I-3 (in 200 mM NaCl), nuclease S1 (Promega) digestion was carried out under conditions such that only the initiation "band 2" complex existed and the saturated "band 1" complex did not. 1 μM of 5′ end ^{32}P-labeled I-3 was incubated in the presence or absence of 1.4 μM g5p in 10 μL of 200 mM NaCl (in 10 mM Tris-HCl, pH 7.4) at 37 °C for 15 min. A nuclease S1 mixture was added to give a final concentration of 0.67 unit/μL of nuclease S1 and 1 mM $ZnCl_2$. The reaction mixture was incubated for 1 min and stopped by adding an equal volume of 88% formamide and 30 mM EDTA. The nuclease-digested samples were briefly heated and then resolved in 12% denaturing polyacrylamide gels (containing 7.6 M urea; acrylamide/bis = 19:1). After being dried, the gel was exposed to a storage phosphor screen. The bands were quantitated, and their positions were assigned according to four parallel reference sequencing lanes of A, G, C, and T.

Since the amount of I-3 complexed with g5p under these conditions accounted for only 10% of the total I-3 (see Results), most signals were contributed by free I-3, and quantitative analysis was needed. Each band of the g5p-bound I-3 and the control lane was quantitated, and the percent protection of the corresponding base by g5p was calculated according to the following equation: 100% × (quantity in band without g5p − quantity in band with g5p)/(quantity in band without g5p). Those sequence positions protected by g5p gave positive values of the percent protection.

Determination of End Boundaries of I-3 in the Initiation Complex. To determine the 3′-end boundary of the region of I-3 needed to form the initiation complex with g5p, I-3 was ^{32}P-labeled at the 5′ end with T4 polynucleotide kinase. Partial nuclease S1 (0.1 unit/μL) digestion was performed at room temperature for 5 min to generate random I-3 fragments. I-3 fragments at a total nucleotide concentration $\approx$ 50 μM were incubated with 0−14 μM g5p at 37 °C for 15 min in TE buffer containing 200 mM NaCl. The mixtures were applied to a membrane filter (0.45 μm, HAWP, containing nitrocellulose; Millipore) to selectively bind protein and protein-containing complexes. The bound DNA was extracted from the filter with phenol/chloroform, ethanol-precipitated, and then resolved in 12% denaturing polyacrylamide gels. The gels were dried and exposed to a storage phosphor screen. Band positions were assigned according to four parallel reference sequencing lanes of A, G, C, and T.

For 5′-end boundary determination, I-3 was labeled at the 3′ end by employing T4 RNA ligase (Promega) and [5′-^{32}P]-pCp (ICN). The same procedure described above was applied except that the range of g5p concentrations was 0−40 μM. Since specific length markers were not available, band positions were assigned by counting each band of the fragment starting from the intact I-3 on a high-resolution phosphor image. This assignment was based on the assumption that each phosphodiester bond of I-3 was accessible to nuclease S1.

Table 1: 36 ssDNA Sequences Recovered after Eight Rounds of SELEX Selection

Clone [a]	Sequence [b]	Base population [c] A	G	C	T
	20 25 30 35 40				
I-3 (19) [d]	-tGGGGTCAGGCTGGGGGTTGTGCAGGTC-	2	14	4	6
G-2 (7) [d]	-tGGGGTCAGGCTGGGGCTGTGCAGGTC-	2	14	5	5
G-8 (2) [d]	-TAGGGCAGGGGTCGTCGGGTTAGGGC-	3	14	4	5
I-4	-tAAGGGCAGGGGTCGTCGGGTTAGGGC-	4	14	4	4
F-4	-tGAGGGCTGGGGTCGTCGGGTTAGGGT-	2	15	3	6
G-1	-tGGGGTCGGCTGCGGGGAGTAGGGGTG-	2	17	3	4
G-13	-CATCGACGGGTGAGGGATAGTCGTCC-	5	10	6	5
F-5	-ATACACTGCTACCTGCCGTCCGGGCC-	4	6	11	5
G-11	-GTACACAGCTACCTGCCGTCCGGGCT-	4	7	10	5
F-9	-CCACACAATCTCCTCACCATTCCCGC-	6	1	14	5
I-7	-GTGCCACCCTCCTCTCTTGTTCTTGT-	1	4	10	11

[a] Clones for sequencing were designated by a letter for the agar plate and a number for the clone from that plate. Where more than one clone had the same sequence, the sequence is named for the first clone. [b] Only the sequences of the N_{26} variable region are shown from 5′ to 3′. Numbering is from the 5′ end of the complete PV-58 sequence. The "t" at the 5′ end of some sequences is from the constant region. Repeated motifs: boxed for CPuGGPy, shaded for TPuGGGPy, and double-underlined for PyPuPuGGGPy (Pu stands for A or G; Py for C or T). [c] Bases in just the variable region. [d] The figures in parentheses refer to the numbers of clones with identical sequences.

RESULTS

ssDNA Sequences Selected using SELEX. SELEX was used to select high-affinity g5p-binding sequences from an ssDNA library of 58-mers, PV-58, that contained a central stretch of 26 nucleotides with approximately random incorporation of the four nucleotides. One nanomole of PV-58 (6×10^{14} sequences) was used for the initial selection. To approximate physiological conditions, the selection was performed at 37 °C in a buffer containing 200 mM NaCl, pH 7.4. After eight rounds, 36 sequences were cloned and sequenced (Table 1). Nineteen clones had an identical sequence (named "I-3"), and seven additional ones differed from I-3 by one base. Although g5p is known to prefer to bind to pyrimidines, 32 of the 36 independently cloned oligonucleotides surprisingly were G-rich. G-centered motifs with three to five purines, CPuGGPy, TPuGGGPy, and PyPuPuGGGPy (Pu stands for A or G, and Py for C or T), repeatedly appeared among most of these sequences (Table 1). The information content (*33*) for the variable region after selection was 25.3 bits, which was an average of 0.97 bit per position. Since the information content of the unselected material averaged only 0.07 bit per position, there was a significant enrichment of sequences during the selection procedure. For example, if each position comprised 80% of a specific base and 20% of equal numbers of the other three bases, the information content would be 0.96 bit per position. For the calculation of information content, all the sequences were aligned as shown in Table 1, were weighted by the number of identical clones, and were corrected for sampling uncertainty (*33*).

The predominant sequence I-3 had two TG_4T and two CAG_2Py motifs, which gave I-3 an unexpected resemblance to a combination of the *Tetrahymena* telomeric repeat (TG_4T; *34*) and the human telomeric repeat (TAG_3T; *35*). I-3 was chemically synthesized for further characterization.

To explore the evolution of the G-rich sequences during SELEX, analyses of the cloned sequences and their base distribution for the intermediate rounds of selection were performed. As shown in Table 2, the variable region of the synthetic PV-58 library was slightly higher in G content than the expected 25%. A preference for guanine in randomly synthesized DNA has been reported elsewhere (*24*). After

Table 2: Base Distributions in the Variable N_{26} Region after Different Numbers of Rounds of SELEX

SELEX round	# of cloned sequences	averaged base population per strand (%)					
		A	G	C	T	Pu	Py
0	16	18.5	32.9	22.4	26.2	51.4	48.6
4	24	16.2	13.1	35.9	34.8	29.3	70.7
6	13	11.2	16.0	41.7	31.1	27.2	72.8
8	36	9.2	49.8	19.2	21.8	59.0	41.0

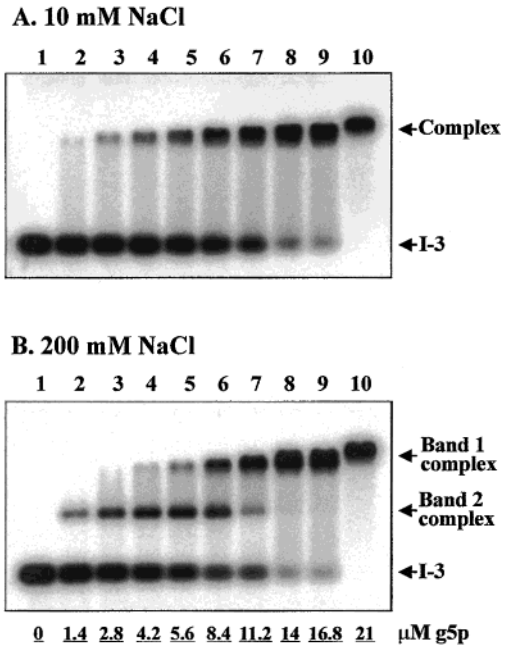

FIGURE 1: EMSA of g5p titrations of I-3 in 10 and 200 mM NaCl. ^{32}P-labeled I-3 (1 μM) was titrated at 37 °C with increasing concentrations of g5p in (A) 10 mM NaCl or (B) 200 mM NaCl. Samples were subjected to 2.5% agarose gel electrophoresis in TAE buffer. The gels were fixed, dried, and analyzed as described in Experimental Procedures. The concentrations of g5p were (from left to right in both panels) 0, 1.4, 2.8, 4.2, 5.6, 8.4, 11.2, 14, 16.8, and 21 μM (per monomer). 14 μM of g5p (lane 8) was a concentration theoretically sufficient to saturate I-3 in the $n = 4$ binding mode. That is, the [protein monomer]/[nucleotide] ratio was P/N = 0.25 in lane 8.

four rounds of selection, each of the 24 cloned sequences was pyrimidine-rich and the G content dropped to 13%. Thus, the early rounds of selection were for moderately high affinity pyrimidine-rich sequences, as expected. Then, the guanine population increased dramatically between the sixth and eighth rounds. The final selected sequences had specific G-rich patterns that evolved from the SELEX procedure and were different from the majority of G-rich sequences in the PV-58 library.

G5p Binds Differently to I-3 at 10 and 200 mM NaCl. In initial experiments to determine the binding affinity of g5p for I-3, I-3 titrations with g5p were analyzed by EMSA. The I-3 (1 μM) was ^{32}P-labeled at the 5′ end and incubated with increasing concentrations (0 to 21 μM) of g5p in 10 and 200 mM NaCl at 37 °C (Figure 1). At 10 mM NaCl, one band of complex appeared (Figure 1A). Since only a minor amount of radioactivity (≤10%) appeared between the complexed and free DNA bands, this binding was approximately all-or-none, reflecting the high cooperativity of g5p. Because one band of complexed I-3 persisted throughout the titration, it was identified as a saturated complex. A slight retardation in the mobility of the saturated complex was always observed in the presence of excess g5p, as shown in lane 10 of Figure 1A. This possibly was due to a well-known switch in binding mode from $n = 4$ to $n = 3$ in the presence of excess g5p (*6*).

Table 3: Apparent Binding Affinities of the g5p Dimer for I-3 and PV-58 in 10 and 200 mM NaCl

	I-3		PV-58	
NaCl	10 mM	200 mM	10 mM	200 mM
$K\omega_{app}$ ($\times 10^{-5}$ M^{-1})	3.0 ± 0.7^a	$>300^b$	2.4 ± 0.2^a	2.1 ± 0.1^a

[a] Values were obtained from EMSA titrations; see Experimental Procedures. Data are the averages of two independent experiments. Errors are the range of values from the two experiments. [b] This is a minimal value and was estimated from the concentration of PV-58 needed to dissociate 50% of the g5p•I-3 band 1 complex in a competition experiment in 200 mM NaCl.

At 200 mM NaCl, the salt concentration at which SELEX selection was performed, we were surprised to observe an additional band (band 2) of complex that was formed at low g5p concentrations (Figure 1B). The band 2 complex appeared prior to the appearance of the saturated (band 1) complex (Figure 1B, lane 2), reached a plateau in the middle of the titration (lanes 4−6), and essentially disappeared before the end of titration (lane 8). The concentration of g5p in lane 8 was sufficient to saturate the I-3 in an $n = 4$ binding mode. Therefore, the band 2 complex appeared to be an intermediate to the formation of a saturated band 1 complex. Since the intermediate complex was not saturated with g5p, we designate it an initiation complex.

Salt Effects on g5p-Binding Affinity of I-3. That the selected DNA sequence I-3 could form different complexes with g5p at different salt concentrations was unusual. The binding affinity of g5p for I-3 was >100-fold higher at 200 mM NaCl than at 10 mM NaCl, whereas the affinity of g5p for PV-58 was slightly reduced at the higher salt concentration (Table 3). These results, combined with the results in Figure 1, suggested that the high affinity of g5p for I-3 was achieved through the formation of the initiation complex at 200 mM NaCl.

EMSA was used to further explore the salt effects on the binding affinity of g5p for I-3. The ^{32}P-labeled I-3 was incubated with g5p and with increasing concentrations of NaCl in the presence or absence of an unlabeled competitor, I-7 (Figure 2). I-7 was the most pyrimidine-rich sequence obtained from the eighth round of SELEX; Table 1. As shown in Figure 2, the competitiveness of I-7 for g5p binding decreased with increasing concentrations of NaCl. Moreover, the initiation complex (band 2) only appeared when the NaCl concentration was ≥ 50 mM. Therefore, the g5p-binding affinity of I-3 was salt-dependent, and the formation of the initiation complex was correlated with higher affinity binding, relative to the pyrimidine-rich sequence I-7, at higher salt concentrations.

CD Titrations of I-3 with g5p at 10 and 200 mM NaCl. CD titrations were used to further characterize the g5p-binding properties of I-3 at 10 and 200 mM NaCl at 37 °C. As shown in Figure 3A, the free I-3 (solid line) in 10 mM NaCl showed a typical CD spectrum of single stranded DNA, and the spectral changes above 245 nm upon addition of g5p were monotonic and were similar to those found for

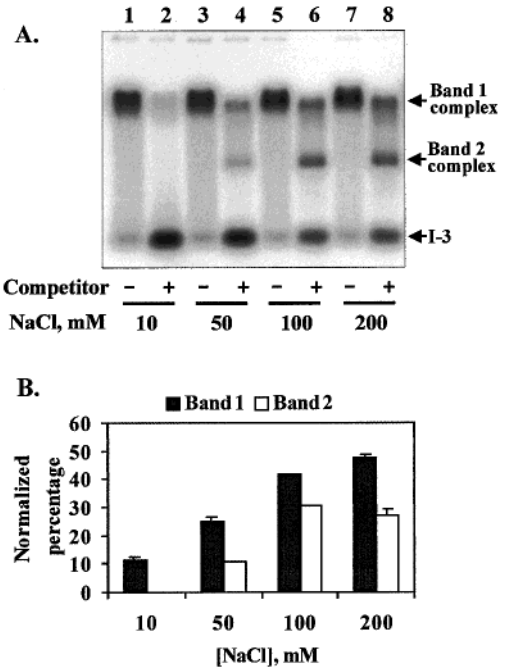

FIGURE 2: Effects of NaCl concentration on g5p affinity for I-3. (A) EMSA of ^{32}P-labeled I-3 (1 μM) that was incubated with 14 μM of g5p at 37 °C for 15 min in the presence (lanes 2, 4, 6, and 8) or absence (lanes 1, 3, 5, and 7) of a competitor (2 μM of I-7). The NaCl concentrations during the incubation were 10, 50, 100, and 200 mM, as shown at the bottom of the figure. Gels were run and analyzed as for Figure 1. The band 2 complex appeared under competitive conditions in which NaCl concentrations were at least 50 mM (lanes 4, 6, and 8). (B) Chart showing the relative percentages of band 1 and band 2 complexes in lanes 2, 4, 6, and 8. The percentages were normalized to lanes 1, 3, 5, and 7, respectively. The data were the averages of two independent experiments, except that only one experiment was performed at 100 mM NaCl. The error bars show the range of values from the two experiments.

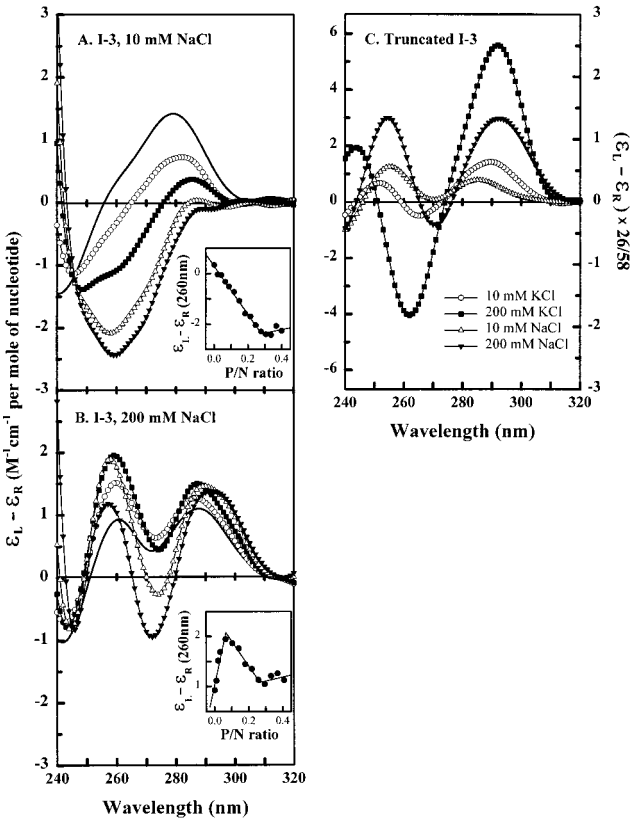

FIGURE 3: CD spectra of the full-length and truncated I-3. The full-length I-3 (about 1 μM strand concentration) was titrated with g5p in 10 mM NaCl (A) or 200 mM NaCl (B) at 37 °C. Representative spectra taken during the titrations are shown. (A) CD spectra at 10 mM NaCl of free I-3 (—) and g5p·I-3 complexes at P/N ratios of 0.07 (○), 0.15 (■), 0.25 (△), and 0.34 (▼). (B) CD spectra at 200 mM NaCl of free I-3 (—) and g5p·I-3 complexes at P/N ratios of 0.02 (○), 0.06 (■), 0.14 (△), and 0.29 (▼). Insets show CD values at 260 nm as a function of P/N ratio. (C) CD spectra of the truncated I-3 in 10 mM KCl (○), 200 mM KCl (■), 10 mM NaCl (△), and 200 mM NaCl (▼) at 37 °C. All spectra are plotted as $\epsilon_L - \epsilon_R$ in units of M^{-1} cm^{-1}, per mole of nucleotide, with values on the left-hand scales. In addition, the right-hand scale for panel C shows nucleotide molar values reduced by a factor of 26/58, which allows a more direct comparison of band magnitudes for the truncated I-3 sequence (panel C) and the full I-3 sequence (panel B) when both are at 200 mM NaCl.

titrations of other ssDNAs with g5p at low salt concentrations (*36*). The end point of this titration was at a P/N ratio of 0.25−0.34 (one g5p monomer per 4 to 3 nucleotides; Figure 3A, inset).

At 200 mM NaCl, CD titrations showed two modes of binding, one at P/N ratios < 0.1 (less than one g5p monomer per 10 nucleotides), followed by the stoichiometric saturation of I-3 at a P/N ratio close to 0.25 (one g5p monomer per 4 nucleotides; Figure 3B, inset). These binding modes were consistent with the results of EMSA (see Figure 1) and with the formation of the initiation complex at low g5p concentrations (smaller P/N ratios), followed by the formation of the saturated complex with increasing concentrations of g5p (larger P/N ratios). In addition, the spectrum of free I-3 in 200 mM NaCl showed two positive bands at about 260 and 290 nm (see Figure 3B, solid line). Since there are four copies of telomeric and telomere-like sequence motifs in the variable region of I-3 (Table 1), I-3 may form an intrastranded G-quartet structure (interstranded interaction was ruled out because I-3 existed in a unimolecular form; see below). Parallel G-quadruplexes have a characteristic CD spectrum with a positive band at 260−265 nm (*37*, *38*), while antiparallel G-quadruplexes show a positive band at 290−295 nm and a negative band close to 260 nm (*37*, *38*). Therefore, unbound I-3 in 200 mM NaCl showed a combination of the CD spectral features of two types of G-quadruplexes. The overall spectrum of free I-3 was most like that previously reported for an *Oxytricha* $(T_4G_4)_2$ hairpin (*39*).

These two positive CD bands were present, but with different intensities, when I-3 was saturated with g5p (see Figure 3B, P/N = 0.29), suggesting that the overall structure of I-3 was not substantially altered by g5p binding.

CD Spectra of a Truncated I-3 Sequence. To further test whether the G-rich central segment could be responsible for the CD spectral features of I-3 in 200 mM NaCl, CD spectra were obtained for a truncated I-3 sequence (5′-GGGGTCAGGCTGGGGTTGTGCAGGTC-3′), denoted I-3c26. This sequence consisted of only the central 26 nucleotides of I-3 and contained 14 G's (see Table 1). CD spectra of I-3c26 displayed large changes as the NaCl concentration was increased from 10 to 200 mM. At 200 mM NaCl, I-3c26 acquired CD bands at about 255 and 292 nm (Figure 3C, solid triangles) that were close to those of the full-length I-3 under the same conditions (Figure 3B, solid line). The magnitudes of the positive CD bands of I-3 and I-3c26 were also in reasonable agreement (ranging from 0.9 to 1.4 M^{-1}

cm^{-1}), if the reduced number of nucleotides in the truncated I-3c26 sequence was taken into account by multiplying the I-3c26 spectrum by a factor of 26/58. (Compare the left-hand scale of Figure 3B with the right-hand scale of Figure 3C). These CD data supported the view that the CD spectrum of the I-3 sequence in 200 mM NaCl was dominated by G-quartets within the central sequence of 26 nucleotides.

The I-3c26 sequence can potentially form a variety of structures containing G-quartets, depending on which G's are involved. It is also known that G-quadruplex structures are influenced by whether the cation is Na^+ or K^+ (*40*, *41*). Intramolecular chair-type structures appear to require the presence of potassium (*41*). As I-3c26 was titrated from 10 to 200 mM KCl, the 292-nm positive band increased, a negative band appeared at 260 nm, and a small positive band appeared above 240 nm (Figure 3C). These features were remarkably similar to those in the CD spectrum of a thrombin binding aptamer in 25 mM KCl (*42*), which is known to fold into an intramolecular chair-form G-quadruplex in which adjacent guanosines along and between strands alternate in their glycosyl (*syn* or *anti*) conformations (*28*, *29*). This provided strong evidence that I-3c26 can indeed form a G-quadruplex structure that is probably an intramolecular chair fold. The specific G-quadruplex fold that might be formed by I-3c26 and the full-length I-3 in the presence of Na^+ ion is not known, but it could conceivably be an intramolecular basket-type fold. The positive 260 nm CD band in the presence of Na^+ could then have its origin in the nonalternating arrangement of glycosyl bonds within the G-tetramers in such a fold (*28*) and/or in a sodium-dependent stacking of nontetrameric G's.

Stoichiometry of the Initiation Complex. Primer-annealing experiments were performed to determine the strand stoichiometry of I-3 in the initiation complex. The method was similar to that used by others to study G-quartet complexes (*43*). ^{32}P-labeled I-3 was incubated with various concentrations of the 16-mer primer that was complementary to the 3′ end of I-3. If I-3 existed as a unimolecular form, one retarded band with I-3 plus an annealed primer, in addition to the primer-free band, should appear during gel electrophoresis. If I-3 existed in an n-stranded form, up to n bands, in addition to the primer-free band, could appear during electrophoresis. Figure 4A, lanes 1−4, shows that only one additional band was detected when the 3′ primer was annealed to free I-3.

Figure 4A, lanes 5−8, further shows that, after the addition of g5p to form an initiation band 2 complex, only one additional band 2 complex appeared upon addition of the 3′ primer. This strongly suggested that free I-3 existed in a unimolecular form in a solution of 200 mM NaCl and that the intermediate with g5p contained one I-3 strand with a free 3′ end. The order of addition of the primer and g5p to I-3 was not important (data not shown), further suggesting that the primer-annealing and g5p-binding sites on I-3 were not overlapping and that about 16 nucleotides at the 3′ end of I-3 were not involved in g5p binding to form the initiation complex. In other experiments, mixtures of I-3 and various concentrations of the 3′ primer were pretreated by heating to 95 °C and slow cooling to room temperature. The results of these experiments (not shown) were essentially identical to lanes 1−4 of Figure 4A. Therefore, it was unlikely that two or more I-3 strands folded asymmetrically so that only one 3′ end was accessible. Finally, gel electrophoresis of serial dilutions of I-3 provided additional evidence that I-3 existed in a unimolecular form at both low and high sodium concentrations. As shown in Figure 4B, when diluted over a 250-fold concentration range of 4 to 0.016 μM in either 10 or 200 mM NaCl, I-3 migrated as a single band in a native gel.

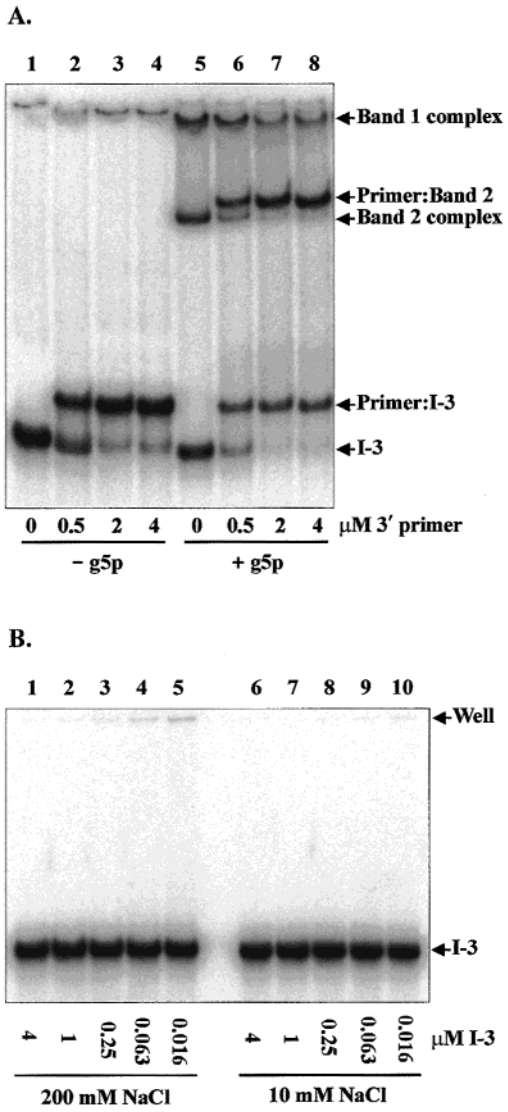

FIGURE 4: Stoichiometry of free and g5p-complexed I-3. (A) Primer annealing of free and g5p-complexed I-3. The 3′ primer (0, 0.5, 2, and 4 μM, as shown at the bottom of the figure) was incubated with 1 μM of ^{32}P-labeled free I-3 (lanes 1−4) or I-3 complexed with 7 μM g5p (lanes 5−8) in 200 mM NaCl buffer. Reaction mixtures were resolved in 12% polyacrylamide gels in TBE buffer. The gels were then fixed, dried, and analyzed. Details are given in Experimental Procedures. (B) Serial dilution of free I-3. Four-fold dilutions of I-3 were made in 200 mM NaCl (lanes 1−5) and 10 mM NaCl (lanes 6−10). A trace amount of ^{32}P-labeled I-3 was added to each dilution. The final I-3 concentrations are shown at the bottom of the figure. Electrophoresis was performed as for panel A.

The molar ratio of protein to DNA in the intermediate band 2 complex was determined by extraction of the band 2 complex, performing SDS−PAGE, and quantitating the amount of stained protein and labeled I-3. Results are tabulated in Table 4. Together with the primer-annealing data, these data revealed that the initiation complex consisted of one I-3 strand and about three g5p dimers (six monomers).

Table 4: Protein/DNA Stoichiometries of the Initiation and Saturation Complexes

	saturation complex (10 mM NaCl)	initiation complex (200 mM NaCl)	saturation complex (200 mM NaCl)
g5p monomer per I-3 strand[a]	15.5 ± 0.6	6.4 ± 1.2	14.8 ± 2.8

[a] Data are shown as mean ± SD from at least three measurements.

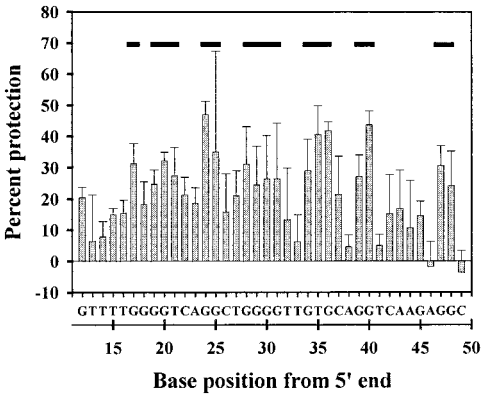

FIGURE 5: Quantitative nuclease S1 footprinting. The 5′ end ^{32}P-labeled I-3 (1 μM) was preincubated with or without 1.4 μM g5p in 200 mM NaCl at 37 °C for 15 min. After treatment with nuclease S1 for 1 min, the mixtures were resolved in 12% denaturing polyacrylamide gels. Quantitative analysis was carried out and the percent protection was plotted as a function of base position of I-3. (See Experimental Procedures for details.) The percent protection is from the average of three independent experiments and standard deviations are shown. The horizontal bars on the top indicate nucleotides that were protected by at least half of the maximum protection (which was 47%).

In contrast, the saturation complex, formed at 10 or 200 mM NaCl, had a stoichiometry that averaged about 15 g5p monomers per I-3 strand (Table 4), consistent with a g5p binding mode of $n = 4$.

The g5p-Binding Sites within the Initiation Complex. The above results showed that g5p did not saturate the whole I-3 sequence within the initiation complex. Nuclease S1 footprinting was used to investigate whether g5p bound directly to the central G-rich region of the selected sequence or whether the g5p bound to and protected the 5′ and 3′ primer ends of some type of folded I-3 structure that was stabilized by the G-rich center. To avoid interference by the band 1 complex, conditions for complex formation were chosen such that the initiation complex was formed but the saturated complex was not formed, i.e., the conditions were the same as in lane 2 of Figure 1B. Results shown in Figure 5 established that the dominant g5p protection was directly within the central G-rich region, extending from nucleotide 17 to 42. With the exception of nucleotide 35, the most highly protected nucleotides were all G's and included two G's in the 3′ region (Figure 5, horizontal bars).

End Boundaries of I-3 in the Initiation Complex. Nuclease S1 footprinting (Figure 5) suggested that the variable region of I-3 contained the actual sites bound by g5p to form the initiation complex. However, protection from nuclease S1 digestion could partially be an indirect effect of g5p binding. Therefore, the sequence boundaries of the variable region that were required for formation of the initiation complex were more exactly defined. To determine the 3′-end boundary, I-3 was ^{32}P-labeled at the 5′ end and partially digested by nuclease S1 to generate fragments of variable lengths that extended from a fixed, labeled 5′ end. When incubated at 200 mM NaCl with decreasing concentrations of g5p, fragments that had reduced g5p-binding affinity were not bound and could not be recovered by filter binding (see Experimental Procedures). The left lane of Figure 6A shows that, as expected, a large range of sequence lengths, including those that were truncated into the variable region, were isolated at saturating concentrations of g5p. At lower concentrations of g5p at which formation of the initiation complex dominated, the lengths of DNA that were isolated (and that could be detected) were longer, representing the loss of successive nucleotides from the 3′ unlabeled end. A discrete boundary was identified, between nucleotides T41 and C42 at 7 μM g5p, or between C42 and A43 at lower g5p concentrations, that represented the 3′ end of the sequence needed to form the initiation complex. This boundary was within a nucleotide of the junction between the variable and 3′ primer region, which was between C42 and A43. The clear identification of this 3′ boundary indicated that G47 and G48, which were relatively protected in the nuclease S1 footprinting experiments (Figure 5), were not part of the actual binding site and were possibly indirectly protected by the bound g5p.

In the reverse experiment, I-3 was labeled at the 3′ end to determine the 5′-end boundary. Figure 6B shows that there was also a discrete 5′ boundary, between T16 and G17 at 3.5−14 μM g5p, or between T16 and T15 at 0.9 μM g5p, where the former was at the junction between the variable and 5′ primer region. Therefore, the boundary experiments confirmed that g5p bound directly to the variable region to form the initiation complex and that the minimum motif for forming an initiation complex with g5p essentially encompassed the entire variable sequence between the junctions with the two primer regions.

DISCUSSION

SELEX of a Cooperative ssDNA Binding Protein. SELEX has been successfully used to isolate, from an ssDNA library of 58-mers (PV-58), those sequences that bind g5p with high-affinity under physiologically relevant conditions (200 mM NaCl, 37 °C, and pH 7.4). The variable region of PV-58 molecules consisted of the central 26 nucleotides. The selection of sequences by g5p was based on the formation of a discrete electrophoretic band of apparently saturated g5p•ssDNA complexes during competitive binding of PV-58 sequences to a limited amount of g5p. Owing to its high cooperativity, g5p will usually bind to and saturate every site on most PV-58 sequences, including the two flanking 16-mer constant sequences. In terms of applying SELEX, this behavior is the major difference between cooperative DNA binding proteins (such as g5p) and specific DNA-binding proteins. The latter type of protein generally binds with one protein molecule per DNA strand, and the binding site is usually selected from and located within the variable region, although part of the primer sequence can be involved in binding (*30*). Given the fact that the binding site for g5p is relatively small (three to four nucleotides per g5p monomer; *6*), the stretch of 26 selectable nucleotides in the central, variable region of PV-58 was apparently long enough to create a structured site that favored the initial binding of

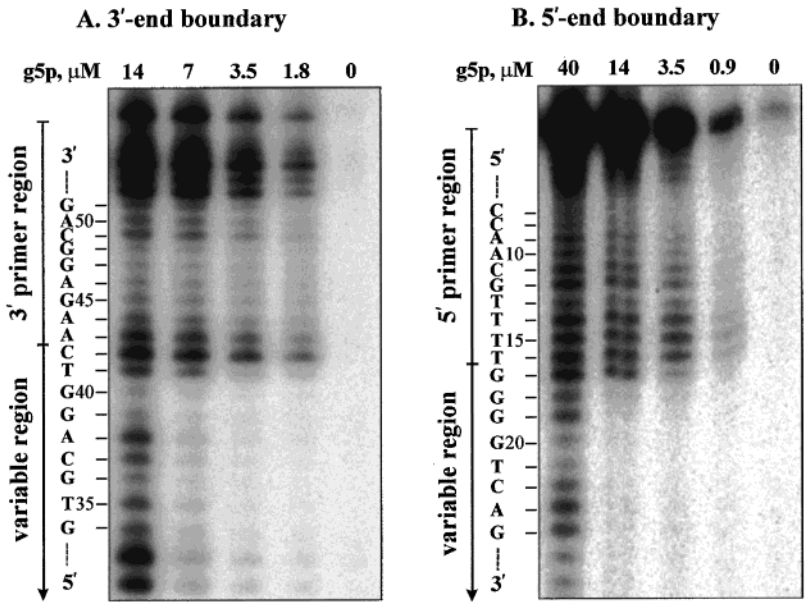

FIGURE 6: End boundaries of the I-3 sequence in the initiation complex with g5p. I-3 was labeled (A) at the 5′ end for determining the 3′-end boundary, and (B) at the 3′ end for determining the 5′-end boundary. Labeled I-3 was partially digested by nuclease S1 and incubated with various concentrations of g5p as shown on the figure. The fragments that could form complexes with g5p were selected by membrane filtration and resolved on 12% denaturing polyacrylamide gels, as described in Experimental Procedures. Bands were assigned, and the variable and two primer regions were identified and are marked on the figure. The nucleotide marking a given band is the last nucleotide (farthest from the labeled end) that is on that fragment.

several g5p dimers. Thus, the cooperative binding nature of g5p, which led to subsequent saturation of the constant sequences, did not prevent the selection of sequences that bound with high affinity.

Characterization of an I-3 Selected Sequence and an Initiation Complex. After eight rounds of selection, most SELEX-derived sequences for g5p binding were G-rich and had one or more similar motifs such as CPuGGPy, TPuGGGPy, and PyPuPuGGGPy (see Table 1). This was surprising because g5p is well-known to bind with higher affinity to synthetic oligonucleotides that are pyrimidine-rich than to those that are purine-rich (*9*, *10*). However, analyses of the cloned sequences from intermediate rounds of selection showed that the selected G-rich sequences were finally selected from larger pools of intermediate sequences that were indeed pyrimidine rich (Table 2). In addition, under the selection conditions of 200 mM NaCl and 37 °C, the g5p-binding affinities for ssDNA sequences that are purine-rich are weakened, because the binding to purine-rich sequences is largely driven by ion release and ionic interactions. Nevertheless, for a predominant selected sequence, I-3, the g5p binding affinity was increased by over 2 orders of magnitude as compared with its averaged binding affinity to the original mixture of PV-58 sequences (Table 3).

Titrations of I-3 with g5p were performed and monitored independently by EMSA (Figure 1) and by CD spectroscopy (Figure 3). The results were consistent in showing that, prior to saturation, an intermediate initiation complex was formed at 200 mM but not at 10 mM NaCl. The formation of this initiation complex appeared to be a key factor that resulted in a higher binding affinity of g5p for I-3 at 200 mM than at 10 mM NaCl (Table 3). CD spectroscopy also suggested that, at 200 mM NaCl, free I-3 formed a structure, probably involving a G-quadruplex, to which, according to the data in Table 4, about three g5p dimers bound to form the initiation complex. The initiation complex was apparently

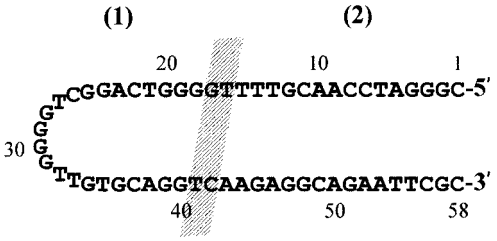

FIGURE 7: Schematic of the alignment of the I-3 sequence when complexed with g5p at 200 mM NaCl. There are two major regions to the structure. Region (1) consists of nucleotides 16−42 and is essentially identical with the selected G-rich variable region. This region may form a structure by folding into a unimolecular G-quadruplex. About three g5p dimers bind directly to this region to form the band 2 initiation complex detected by EMSA. The shaded bar highlights the nucleotides that are at the 3′- and 5′-end boundaries. Region (2) consists of the 5′ and 3′ primer sequences. These presumably are oriented in an antiparallel fashion and provide binding sites that are subsequently saturated to form the band 1 complex in EMSA experiments.

formed in an all-or-none fashion, because there were no intermediate bands between the free I-3 and band 2 on agarose gels (Figure 1). That is, it appeared that all three g5p simultaneously bound to form a core initiation complex for further saturation. Moreover, CD spectroscopy showed that the I-3 structure was maintained largely intact within the saturated complex.

The binding site for the g5p dimers that form an initiation complex was essentially identical to the 26-mer, central G-rich selected region of I-3 (Figures 3, 5, and 6). Our knowledge about the initiation structure is illustrated in Figure 7. One I-3 molecule is involved. The central 26-mer (nucleotides 17−42) is G-rich, and this sequence has the potential to be folded into one of several unimolecular G-quadruplexes, such as the chair form described for the thrombin aptamer in the presence of K^+ (*28*, *29*). Although the structure of I-3 in the presence of Na^+ is not known, it

is not likely to be a simple chair G-quadruplex (Figure 3). Three g5p dimers bind directly to this central region to form an initiation complex that migrates as "band 2" in EMSA experiments. We assume that the 3′ and 5′ primer ends are juxtaposed in an antiparallel fashion for subsequent saturation by additional g5p dimers (which have rotationally symmetric binding sites) to form the "band 1" complex. A key point is that the putative G-quadruplex structure may not only stabilize a desirable template for binding, but the structure itself appears to be the actual initial g5p binding site.

Biological Relevance. The abilities of Ff g5p to form both a cooperatively saturated complex through nonspecific binding and an unsaturated complex through specific binding correspond to the biological functions of g5p in sequestering the viral genome and regulating the translation of viral mRNAs, respectively. The g5p is known to saturate the nascent Ff ssDNA genome for phage genome packaging (*1*). In this sense, g5p acts as a non-sequence-specific ssDNA binding protein and binds in a cooperative manner, even though the nucleation process is still unknown. On the other hand, g5p is also involved in translational regulation of some viral mRNAs, such as gene 2 mRNA, by binding to a specific sequence in the 5′-end untranslated region of the mRNA (see below). The behavior of g5p binding to the SELEX-derived sequences can provide insight into how cooperative interactions are initiated under physiological conditions and into the types of sequences and/or structures of DNA or RNA that might be specific binding sites.

In addition to the in vitro-selected I-3 sequence, a naturally occurring sequence has also been found to form an unsaturated intermediate complex with g5p. Michel and Zinder (*17*) showed that the first 16 nucleotides (5′-GUUUUUGGGGC-UUUUC-3′) of the Ff gene 2 mRNA leader sequence is required for g5p-mediated translational repression of this mRNA and that an RNA (208−211 bases in length) containing this leader sequence forms an intermediate complex in gel electrophoresis before a saturated complex is formed (*17*, *18*). The g5p binds to this RNA with a 10-fold higher affinity than to the control RNA. How g5p binds specifically to this sequence is still not clear, but the apparent lack of structure in this region was originally thought to be a dominant factor (*14*). However, a likely tetraplex structure with a central block of G-quartets has recently been shown to form with four strands of the gene 2 leader sequence or with its DNA analogue (*44*). Kneale and co-workers also demonstrated preferential binding of g5p to this structure (*45*). They propose that tails of four antiparallel 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 initiate g5p binding (*45*). In their model, g5p does not bind directly to the G-quartet structure.

Our data with I-3 provides support for the idea that g5p prefers to bind to a structured sequence, but in a somewhat different manner than that proposed by Oliver et al. (*45*) for the gene 2 leader sequence. The I-3 structure is formed with only one strand (Figure 4). Moreover, our data show that G_4 blocks play a direct role in the high affinity binding (Figures 5 and 6). The conformation of I-3 in the initiation complex presumably orients two antiparallel strands (the 5′ and 3′ primer strands; see Figure 7) to form additional g5p dimer binding sites that are subsequently saturated; however, these are not the first sites to be bound by g5p, as might be expected by analogy with the structure of the gene 2 leader sequence.

It remains to be seen whether the $(g5p)_3$•I-3 initiation complex has a biological counterpart or whether such a complex would be formed with RNA. It is relevant to point out that an intrastrand interaction could occur between the gene 2 leader 16-mer and nearby sequences of the gene 2 mRNA to form a secondary structure. Specifically, in the gene 2 leader sequence there is another G_4 block (part of the Shine-Dalgarno sequence) that is downstream and separated from the G_4 block in the leader 16-mer by only 18 bases (*46*). Interaction of these two G_4 blocks through antiparallel G:G pairing could plausibly form a structure to facilitate g5p binding. On the other hand, we found that the specific SELEX-selected motifs (such as CAGGPy and NGGGN; see Table 1) occurred less frequently in the Ff genome (60 times) than in a random sequence of 6000 nucleotides (88 times) and that the motifs did not have an unusual distribution. Therefore, the biological relevance of these specific motifs within the Ff genome remains uncertain.

Protein Binding to G-rich Sequences. G-quartets are also found in other SELEX-derived aptamers for thrombin (*27*), elastase (*47*), and IgE (*48*). G-rich telomere sequences are bound by α and β telomere-binding proteins and by the Cdc13p telomerase-loading protein (*39*, *49*−*51*). Human DNA topoisomerases I and II interact with G-quartet structures (*52*, *53*) and other proteins that interact with G-quartets have been reviewed (*52*, *54*). Loops on intrastrand G-quartet or hairpin structures can be the sites for target binding (*22*). In the case of I-3, some of the loop nucleotides may just be responsible for connecting g5p-binding sites and positioning them in an appropriate three-dimensional conformation, since they appear not to be protected by g5p from nuclease digestion (Figure 5). Nevertheless, the many examples of proteins that bind to G-rich sequences suggest that G-rich motifs selected by the g5p may have features that are recognized by other proteins, and the study of g5p•ssDNA initiation complexes could be of general importance for understanding the binding, or initiation of cooperative binding, of other single-stranded DNA-binding proteins.

ACKNOWLEDGMENT

We are grateful to Dr. Tung-Chung Mou (University of Texas at Dallas) for generously providing purified wild-type g5p and to Dr. Andy Peek of Cytoclonal Pharmaceutics, Inc. (Dallas) for generating a random DNA sequence. We have appreciated the advice and encouragement of Drs. Larry Gold (Department of Molecular, Cellular, and Developmental Biology, University of Colorado, Boulder, CO), Dennis L. Miller (University of Texas at Dallas), and Thomas C. Terwilliger (Los Alamos National Laboratory, Los Alamos, NM) throughout the course of this work.

REFERENCES

1. Model, P., and Russel, M. (1988) in *The Bacteriophages* (Calendar, R., Ed.) pp 386−390, Plenum Press, New York.
2. Terwilliger, T. C. (1996) *Biochemistry 35*, 16652−16664.
3. Alberts, B., Frey, L., and Delius, H. (1972) *J. Mol. Biol. 68*, 139−152.
4. Gray, C. W. (1989) *J. Mol. Biol. 208*, 57−64

5. Skinner, M. M., Zhang, H., Leshnitzer, D. H., Guan, Y., Bellamy, H., Sweet, R. A., Gray C. W., Konings, R. N. H., Wang, A. H.-J., and Terwilliger, T. C. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 2071−2075.
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. Kowalczykowski, S. C., Bear, D. G., and von Hippel, P. H. (1981) in *The Enzymes* (Boyer, P. D., Ed.) pp 373−444, Academic Press, New York.
9. Bulsink, H., Harmsen, B. J., and Hilbers, C. W. (1985) *J. Biomol. Struct. Dyn. 3*, 227−247.
10. Mou, T. C., Gray, C. W., and Gray, D. M. (1999) *Biophys. J. 76*, 1537−1551.
11. Bauer, M., and Smith, G. P. (1988) *Virology 167*, 166−175.
12. Webster, R. E., Grant, R. A., and Hamilton, L. A. (1981) *J. Mol. Biol. 152*, 357−374.
13. Fulford, W., and Model, P. (1988) *J. Mol. Biol. 203*, 39−48.
14. Zaman, G., Smetsers, A., Kaan, A., Schoenmakers, J., and Konings, R. (1991) *Biochim. Biophys. Acta 1089*, 183−192.
15. Model, P., McGill, C., Mazur, B., and Fulford, W. D. (1982) *Cell 29*, 329−335.
16. Yen, T. S., and Webster, R. E. (1982) *Cell 29*, 337−345.
17. Michel, B., and Zinder, N. D. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 4002−4006.
18. Michel, B., and Zinder, N. D. (1989) *Nucleic Acids Res. 17*, 7333−7344.
19. Cheng, X., Harms, A. C., Goudreau, P. N., Terwilliger, T. C., and Smith, R. D. (1996) *Proc. Natl. Acad. Sci. U.S.A. 93*, 7022−7027.
20. Tuerk, C., and Gold, L. (1990) *Science 249*, 505−510.
21. Ellington, A. D., and Szostak, J. W. (1990) *Nature 346*, 818−822.
22. Gold, L., Polisky, B., Uhlenbeck, O., and Yarus, M. (1995) *Annu. Rev. Biochem. 64*, 763−797.
23. Louhan, C. T., and Szostak, J. W. (1995) *J. Am. Chem. Soc. 117*, 1246−1257.
24. Bianchi, A., Stansel, R. M., Fairall, L., Griffith, J. D., Rhodes, D., and de Lange, T. (1999) *EMBO J. 18*, 5735−5744.
25. Hermann, T., and Patel, D. J. (2000) *Science 287*, 820−825.
26. Wilson, D. S., and Szostak, J. W. (1999) *Ann. Rev. Biochem. 68*, 611−647.
27. Bock, L. C., Griffin, L. C., Latham, J. A., Vermaas, E. H., and Toole, J. J. (1992) *Nature 355*, 564−566.
28. 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.
29. Wang, K. Y., McCurdy, S., Shea, R. G., Swaminathan, S., and Bolton, P. H. (1993) *Biochemistry 32*, 1899−1904.
30. Schneider, D. J., Feigon, J., Hostomsky, Z., and Gold, L. (1995) *Biochemistry 34*, 9599−9610.
31. Laemmli, U. K. (1970) *Nature 227*, 680−685.
32. Steinberg, T. H., Jones, L. J., Haugland, R. P., and Singer, V. L. (1996) *Anal. Biochem. 239*, 223−237.
33. Schneider, T. D., Stormo, G. D., and Gold, L. (1986) *J. Mol. Biol. 188*, 415−431.
34. Blackburn, E. H., and Gall, J. G. (1978) *J. Mol. Biol. 120*, 33−53.
35. Moyzis, R. K., Buckingham, J. M., Cram, L. S., Dani, M., Deaven, L. L., Jones, M. D., Meyne, J., Ratliff, R. L., and Wu, J. R. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 6622−6626.
36. Mark, B. L., Terwilliger, T. C., Vaughan, M. R., and Gray, D. M. (1995) *Biochemistry 34*, 12854−12865.
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. Laporte, L., Benevides, J. M., and Thomas, G. J., Jr. (1999) *Biochemistry 38*, 582−588.
40. Hardin, C. C., Henderson, E., Watson, T., and Prosser, J. K. (1991) *Biochemistry 30*, 4460−4472.
41. Marathias, V. M., and Bolton, P. H. (2000) *Nucleic Acids Res. 28*, 1969−1977.
42. Smirnov, I., and Shafer, R. H. (2000) *Biochemistry 39*, 1462−1468.
43. Venczel, E. A., and Sen, D. (1993) *Biochemistry 32*, 6220−6228.
44. Oliver, A. W., and Kneale, G. G. (1999) *Biochem. J. 339*, 525−531.
45. Oliver, A. W., Bogdarina, I., Schroeder, E., Taylor, I. A., and Kneale, G. G. (2000) *J. Mol. Biol. 301*, 575−584.
46. Beck, E., and Zink, B. (1981) *Gene 16*, 35−58.
47. Lin, Y., Padmapriya, A., Morden, K. M., and Jayasena, S. D. (1995) *Proc. Natl. Acad. Sci. U.S.A. 92*, 11044−11048.
48. Wiegand, T. W., Williams, P. B., Dreskin, S. C., Jouvin, M. H., Kinet, J. P., and Tasset, D. (1996) *J. Immunol. 157*, 221−230.
49. Fang, G., Gray, J. T., and Cech, T. R. (1993) *Genes Devel. 7*, 870−882.
50. Hemann, M. T., and Greider, C. W. (1999) *Nucleic Acids Res. 27*, 3964−3969.
51. Nugent, C. I., Hughes, T. R., Lue, N. F., and Lundblad,V. (1996) *Science 274*, 249−252.
52. Arimondo, P. B., Riou, J.-F., Mergny, J.-L., Tazi, J., Sun, J.-S., Garestier, T., and Hélène, C. (2000) *Nucleic Acids Res. 28*, 4832−4838.
53. Chung, I. K., Mehta, V. B., Spitzner, J. R., and Muller, M. T. (1992) *Nucleic Acids Res. 20*, 1973−1977.
54. Williamson, J. R. (1994) *Annu. Rev. Biophys. Biomol. Struct. 23*, 703−730.

BI010109Z