Evidence for Sequence-Specific Recognition of DNA by Anti-Single-Stranded DNA Autoantibodies†

Shawn Y. Stevens and Gary D. Glick*

Department of Chemistry, University of Michigan, Ann Arbor, Michigan 48109-1055

Received August 7, 1998; Revised Manuscript Received October 22, 1998

ABSTRACT: Anti-DNA autoantibodies are a serological hallmark of the autoimmune disorder systemic lupus erythematosus. In a process involving antigen recognition, these antibodies are also believed to mediate the kidney inflammation that results in much of the morbidity and mortality associated with lupus. However, the nature of specific DNA antigens recognized by anti-DNA and the way in which anti-DNA interact with these molecules remains poorly understood. As a first step in defining the molecular basis of anti-DNA interactions, binding site selection experiments were conducted using three clonally related murine monoclonal anti-ssDNA autoantibodies previously isolated from a lupus prone MRL-*lpr* mouse (Swanson, P. C., Ackroyd, P. C., and Glick, G. D. (1996) *Biochemistry 35*, 1624−1633). Studying the interaction of these autoantibodies with the selected sequences (and variants) through affinity measurements and footprinting experiments provides evidence for sequence-specific binding of ssDNA. However, despite the similarity in amino acid sequence between the three mAbs, only mAb 11F8 appears to possess sequence specificity. The salient features of the interaction between 11F8 and its selected sequence (e.g., limited dependence of ionic strength upon binding, cross-reactivity, and conformational complementarity) are best described by combining the paradigms invoked to explain protein·nucleic acid and antibody·antigen recognition.

The interaction between an antibody and antigen in many ways epitomizes specific, high-affinity biomolecular recognition. Although a basal level of binding can be attained through germline-encoded antibodies (*1*), the high degree of specificity observed in a typical immune response results from somatic hypermutations that arise under the selective pressure of an antigen (*2*). The capacity of antibodies to discriminate between an extremely diverse range of foreign entities and self is one of the most intriguing and physiologically important examples of biological specificity. Loss of this self-tolerance can lead to one of several disabling, and sometimes fatal, autoimmune disorders.

Systemic lupus erythematosus (SLE)[1] is a multisystem disorder of unknown etiology that is the prototype of human autoimmune diseases. SLE is characterized by the spontaneous production of autoantibodies against a variety of nuclear and cellular antigens, including DNA (*3*, *4*). A subset of the autoantibodies that bind DNA (both ssDNA and dsDNA) mediates an inflammatory response that often leads to kidney damage (known as lupus nephritis, a form of glomerulonephritis) in SLE patients (*5*−*7*). This process may begin by either the deposition of circulating anti-DNA·DNA immune complexes within glomeruli or the formation of these complexes within the kidney itself on the GBM (*8*). Although these two processes differ somewhat, each requires the same first step: binding of a pathogenic anti-DNA to its antigen. Hence, determining the ligand binding properties of anti-DNA is a necessary step in elucidating the mechanism of renal damage in SLE. On a more general level, this information will also provide insight about how immunoglobulins effect molecular recognition of single- and double-stranded DNA ligands.

At present, little is known about the factors governing anti-DNA·DNA binding and recognition. In particular, it is not yet known whether lupus anti-DNA are sequence-specific and, if so, how specific sites are selected from all others. Here we report the results of binding-site selection experiments designed to determine if mAbs 11F8, 9F11, and 15B10, which are representative of the anti-ssDNA expressed in lupus-prone mice (*9*, *10*), are sequence-specific. To the best of our knowledge, these experiments represent the first application of in vitro selection technology to explore the sequence specificity of anti-ssDNA in particular and ssDNA binding proteins in general. Despite the similarity in amino acid sequence between mAbs 11F8, 9F11, and 15B10, detailed characterization of the interaction between the in vitro selected sequences and the mAbs indicates that only 11F8 is sequence-specific (*16*−*18*). Moreover, the salient features of the interaction between 11F8 and its selected sequence (e.g., limited dependence of ionic strength upon binding, cross-reactivity, and conformational complementarity) are best described by utilizing the paradigms developed to explain protein·nucleic acid and antibody·antigen recognition.

MATERIALS AND METHODS

Nucleic Acid Synthesis and Purification. Solid-phase DNA synthesis was conducted on a Millipore Expedite 8909 DNA/

† Supported by NIH Grant GM 42168.

* Phone: (734)764-4548. Fax: (734)763-2307. E-mail: gglick@umich.edu.

[1] anti-DNA autoantibodies, anti-DNA; double-stranded DNA, dsDNA; ethylnitrosourea, ENU; glomerular basement membrane, GBM; monoclonal antibodies, mAbs; single-stranded DNA, ssDNA; systemic lupus erythematosus, SLE.

10.1021/bi981899o CCC: $18.00
Published on Web 12/22/1998

RNA synthesizer using standard protocols. For in vitro selection and protein binding studies, oligonucleotides were deprotected with NH_4OH (1 mL, 8 h, 55 °C) and purified with oligonucleotide purification cartridges (Applied Biosystems or Millipore). Oligonucleotides used for NMR experiments were purified by reverse-phase HPLC using a Waters Associates chromatograph and absorbance detector (260 nm) for peak detection. Purification of other oligonucleotide samples was conducted using Vydac C4 columns (4.6 × 250 mm for analysis, 1 mL/min; 7.9 × 250 mm for preparative runs, 4 mL/min) eluting with triethylammonium acetate (0.05 M, pH 6.6) and CH_3CN as the mobile phases. All sequences were exchanged into the sodium form prior to analysis by incubation in NaCl (4 M) overnight at 4 °C followed by desalting using SepPak columns (Millipore).

Representative Selection Protocols. MAbs were isolated and purified as previously described (*9*). All of the polymerase chain reaction (PCR) amplification was conducted with *vent* polymerase. Selections were conducted by incubating the mAbs and DNA in 1X TBE (89 mM Tris, 89 mM boric acid, and 1.3 mM EDTA; pH 8.3) containing NaCl (145 mM) at room temperature for 45 min. The first cycle contained equimolar amounts of pool DNA and protein (0.05 nM); dA_{21} (~1 μM) was added as a nonspecific competitor. In subsequent cycles, the selected and amplified DNA pools were added in excess over the [mAb] to increase the selection pressure (cycle 2, 10×; cycles ≥3, 25×). After the fourth cycle in each selection, a high-affinity competitor (dT_{21}) was added in place of dA_{21}. This competitor was incubated with the mAbs for 30 min before pool DNA was added. For mAb 9F11, the incubation time was increased to 2.5 h for the last two cycles with no dT_{21} preincubation after cycle 7. The selections were judged complete when improvements in affinity were no longer observed (nine rounds).

After the incubation to equilibrate the mAb·DNA complexes, the bound and free DNA were separated by two methods. For the first three cycles, samples were filtered with a gentle vacuum through HAWP type nitrocellulose filters (Millipore). The filters were washed with 1X TBE (300 μL) and then swirled in 1X TBE (5 mL) to remove residual unbound DNA. Bound DNA was isolated by incubating the filter in guanidine thiocyanate (350 μL of 4 M solution) at 95 °C for 5 min. The samples were phenol-extracted and ethanol-precipitated with glycogen as a carrier. After the third cycle of selection and amplification, the mAb·DNA complexes were electrophoresed at 4 °C on nondenaturing gels [(5%, 5.1:0.9 acrylamide/bis(acrylamide)] at 300 V for 2.5 min and then 106 V for 30 min. The bound DNA was excised and eluted by soaking in guanidine thiocyanate (350 μL of a 4 M solution) at room temperature overnight. The samples were isolated by phenol extraction as described above.

The selected DNA obtained after separation was amplified by PCR using the minimum number of cycles required to obtain adequate product. The full-length constructs and primers used in the various selection experiments are given in Chart 1, where the asterisk denotes a 5′ biotin tag. For selections 2 and 3 constructs, $n = 7$ and 11, respectively. Note that for a given selection experiment the same set of primers was used for each individual cycle of selection and amplification. The PCR reactions were ethanol-precipitated, loaded onto a denaturing polyacrylamide sequencing gel (8%, 19:1 acrylamide/bis(acrylamide), 8 M urea), and electrophoresed at 62 W for 2 h. The gel was stained with ethidium bromide, and the unbiotinylated strand was excised under long-wave ultraviolet light. The DNA was eluted from the gel overnight in water (1 mL), extracted with *n*-butanol to remove the ethidium bromide, and ethanol-precipitated. Prior to the beginning of the next cycle, the DNA was quantified with the ethidium bromide spot fluorescence method (*11*). Single-stranded DNA from the final cycle of each selection was sequenced by either using the TA Cloning Kit (Invitrogen) or cloning into pUC19 and sequencing colonies with Sequenase 2.0 (United States Biochemicals).

Chart 1

Selection 1:

primers (top) and full length construct (bottom).

5′ GGCAAGGAGCAAGCGTCT and
5′ *GAGTGAAGTGGCGCGTTGTCG

5′ GGCAAGGAGCAAGCGTCTAAN$_7$-
AACGTCGACAACGCGCCACTTCACTC

Selections 2 and 3:

primers (top) and full length constructs (bottom):

5′ GGCAAGGAGCAAGCGTCT and
5′ *TTTGCTGGGCAGCTTTGTCAC

5′ GGCAAGGAGCAAGCGTCTCGTGCTG(N)$_n$-
TTGTGACAAAGCTGCCCAGCAAA

Dissociation Constant Measurements. The affinity of mAb·DNA complexes was measured using a gel shift assay employing ^{32}P end-labeled DNA. For these experiments, mAb·DNA complexes were incubated in 1X TBE with NaCl (145 mM), bovine serum albumin (50 μg/mL), and 5% sucrose (w/v) at room temperature for 45 min. Bound and free DNA were separated by electrophoresis on MetaPhor XR agarose gels (3%) for 30 min at 100 V. The gels were dried and then exposed to a phosphor screen overnight, and the bands were quantified using a Molecular Dynamics Storm 840 phosphorimager with the ImageQuant software. The data were fit via nonlinear least-squares regression to the single-site binding isotherm as previously described (*12*). Errors to the curve fits in no case yielded R^2 values < 0.97. Each affinity measurement was conducted at least three times, affording standard deviations ≤ 10% of the average apparent dissociation constant (K_d).

Footprinting DNA Ligands and mAb·DNA Complexes. MAb and 5′ ^{32}P end-labeled DNA were incubated in buffer (10 mM Tris, pH 8.0; 5 mM NaCl or 150 mM NaCl) containing dA_{21} as a nonspecific competitor (1 μM). In these experiments, excess mAb was used to ensure that all DNA was bound. Potassium permanganate and ethylation interference footprinting were then performed as previously described (*13*). For ligand mapping, potassium permanganate footprinting experiments performed in the absence of protein were conducted following the same protocol. The chemically modified DNA isolated from the footprinting reactions was electrophoresed on a denaturing gel (12% 19:1 acrylamide/bis(acrylamide), 8 M urea polyacrylamide gel in 1X TBE) at a constant power of 62 W for 2 h. The gels were autoradiographed overnight and quantified using a Molecular Dynamics laser densitometer with the ImageQuant software, and the data were analyzed as previously described (*13*, *14*).

Cytosine Mapping of DNA Ligands. 5′ ^{32}P end-labeled DNA was incubated in potassium monopersulfate (240 μM) and KBr (480 μM) for 30 min at room temperature, quenched with addition of HEPES (20 mM), and ethanol-precipitated (*15*). The DNA was dissolved in piperidine (10% in water) and incubated at 94 °C for 30 min. The samples were dried under vacuum, rinsed with water, and dried again. The pellet was dissolved in loading buffer (80% formamide, 10 mM NaOH), and an equal number of cpm was loaded in each pair of lanes. The DNA was electrophoresed on a denaturing gel (12% 19:1, 8 M urea polyacrylamide gel in 1X TBE) at 62 W, constant power, for 2 h. The data were analyzed on a Molecular Dynamics Personal Densitometer running the ImageQuant software.

UV Thermal Denaturation Studies. DNA samples were each diluted in phosphate buffer (10 mM sodium phosphate, 1 mM EDTA, and 145 mM NaCl; pH 7.0), degassed with argon, and sealed in a 1 cm long path length cuvette. The samples were denatured and annealed by heating to 95 °C at a rate of 5 °C/min; they were then cooled to 10 °C at a rate of 3 °C/min using a Cary 3 spectrophotometer equipped with a Peltier temperature controller. After equilibration at this temperature for 30 min, the samples were heated at 1 °C/min. Measurements were taken twice per minute at 260 nm with a bandwidth of 2 nm. T_m values were calculated using the instrument software.

RESULTS AND DISCUSSION

Several factors contribute to affinity and specificity in complexes between proteins and nucleic acids (*16*, *17*). Hydrogen bonding has traditionally been thought to provide the dominant source of specificity (*16*−*19*), with shape complementarity maximizing van der Waals interactions, thus helping to align specific electrostatic and hydrophobic groups (*18*). Ligands that easily change conformation can bind to a larger number of targets than rigid molecules (*19*). However, one consequence of flexibility may be decreased specificity and affinity resulting from the energetic cost of "induced fit". Therefore, an "ideal" complex involves binding a ligand with a well-defined conformation that maximizes specific interactions with the protein and minimizes the loss of entropy upon binding (*19*).

In vitro selection is a powerful method that facilitates rapid screening of nucleic acid libraries for molecules possessing specific properties (*20*, *21*). One variant of this method, binding-site selection, has been used to assess the sequence specificity of both dsDNA (*22*−*24*) and RNA binding proteins (*25*, *27*). In a binding-site selection experiment, a contiguous stretch of single- or double-stranded nucleic acid corresponding to the binding site is randomized. Repetitive cycles of selection and amplification should afford a unique sequence (or small group of sequences) for proteins that are sequence-specific, and a larger number of sequences should result for proteins that are less specific. We used binding-site selection experiments to determine if three clonally related anti-DNA mAbs, 11F8, 9F11, and 15B10, are sequence-specific. These mAbs are part of a panel of anti-DNA previously isolated from a lupus-prone MRL-*lpr* mouse (*9*), a well-established model of human lupus (*26*).

The general binding properties of 11F8, 9F11, and 15B10 are similar to those of other lupus anti-ssDNA: they each bind oligo(dT) > oligo(dG) and exhibit little or no affinity for oligo(dC) or oligo(dA) (*9*). Fluorescence-quenching measurements indicate that the binding-site sizes of mAbs 11F8, 9F11, and 15B10 are about five bases long (*9*). Hence, to sample sequence space, we designed an in vitro selection construct possessing a seven base long random region. We chose to randomize seven rather than five bases because the fluorescence data are approximate and may underestimate the true binding-site size. The random 7-mer was flanked on both sides by DNA of known sequence (the "constant" region) which was designed to prohibit formation of a stable secondary structure because the initial goal of these experiments was to investigate sequence specificity in the absence of a defined structure. Within the constant region were primers for PCR amplification of the selected sequences (see Materials and Methods). It is important to note that because the aim of our experiments was not to identify aptamers for our mAbs but rather to address sequence specificity, constructs with large random inserts (e.g., 30 bases long) which would necessarily sample both sequence *and* conformational space were avoided.

The first mAb we studied was 11F8, which is the only mAb in our panel that binds in vitro to DNA adherent to the GBM, a primary site of renal damage within kidney tissue (*28*). This interaction is thought to be a first step in the progression of events that can lead to kidney damage in SLE (*5*−*8*). After nine cycles of selection and amplification with 11F8 using the selection 1 construct (see Materials and Methods), the selected pool contained two sequences, 5′-GGTCCTT and 5′-GCCTTCT. The selected DNAs migrated faster than a 53 base long DNA size marker on nondenaturing polyacrylamide gels, which suggests that they may possess secondary structure. Because the random region itself is too small to fold into a complex structure and the constant region was designed not to possess secondary structure, this observation was unexpected.

Analysis of the sequences flanking the seven selected bases revealed single consistent mutations *within the constant region*: 5′-AAN$_7$AA changed to 5′-AAGGTCCTTA*C* (**1**) or 5′-A_GCCTTCTAA (**2**), where a dA was deleted. Removing these mutations from the full length sequences in synthetic oligonucleotides restored the mobility to that of a linear 53 base long marker. Inspection of the two selected sequences suggested that **1** and **2** can form hairpins with a four base long loop consisting of 5′-CCTT separated from an internal loop or bulge, respectively, by three base pairs (Figure 1). NMR measurements of an oligonucleotide containing the central 25 bases of **1** are also consistent with the secondary structure shown in Figure 1.[2] Taken together, these observations suggest that mutations produced in the PCR amplification can lead to stable secondary structures.

To ensure that the constant region sequence did not bias the outcome of this first selection in favor of stem−loop structures, additional experiments were performed with random regions of 7 and 11 bases flanked by a constant region different from the selection 1 sequence (selections 2 and 3 constructs, respectively; see Materials and Methods). Inspection of the ligands generated from these new selections suggests that they also form hairpin structures in which a

[2] See Supporting Information.

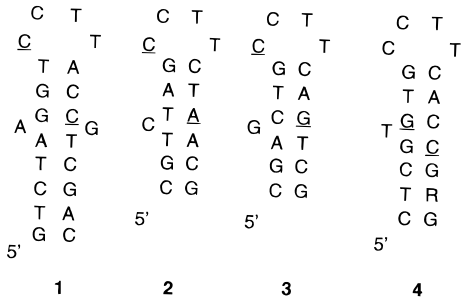
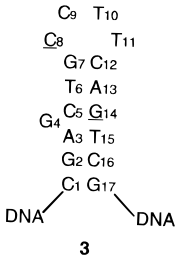


FIGURE 1: Proposed recognition motifs for 11F8 (the selected region in each sequence is underlined). Ligands **1** and **2** are derived from a seven base long random region using the selection 1 construct (see Materials and Methods). Sequences **3** and **4** came from 7 and 11 base long random regions, respectively, using the constructs and primers designated selections 2 and 3. Only the selected regions and regions believed to be base-paired are shown.

5′-CCTT loop is separated from a bulge by three base pairs (**3** and **4**; Figure 1). Although the bulge composition resulting



from each selection differs, the apparent K_d values of **1**−**4** with 11F8 are within 3-fold. This finding suggests that 11F8 does not interact directly with the bulge base itself. Because **1**−**4** bind 11F8 with similar affinity, sequence **3** was chosen for the majority of further experiments as it came from the only selection that resulted in a single sequence (as judged both by sequencing the final pool directly and by cloning the final pool and sequencing a number of clones).

A series of spectroscopic and footprinting experiments were conducted to investigate the structure of **3**. Introducing mutations into the proposed stem region of **3** decreases the T_m values to an extent expected for duplex mismatches (*29*−*31*) whereas a compensatory mutation (dT6·dA13 to dA6·dT13; **M11**) does not impart a significant change in stability (Table 1). In addition, eliminating the bulge increases the T_m by ∼12 °C, which is comparable with other examples of single-base bulges in duplex DNA (*30*, *31*). Potassium permanganate selectively oxidizes the C5−C6 double bond of thymine residues that are not Watson−Crick hydrogen-bonded (*32*). When base pairing and base stacking are eliminated at 94 °C, approximately equal reactions with $KMnO_4$ are observed for all thymine bases in **3**. In contrast, little modification is observed in the proposed hairpin stem at 25 °C (Figure 2).

Unpaired cytosines within a duplex can be probed by reaction with monopersulfate/KBr which selectively brominates the C5−C6 double bond (*15*). We used this reagent to study **2** because both the loop and bulge can be probed. Four dC residues in **2** exhibit enhanced cleavage after bromination: the two loop cytosines, the putative bulge dC, and the cytosine at the stem−loop junction (data not shown). Collectively, these data support the hypothesis that the selected ligands for 11F8 are hairpins with a four base long

Table 1: Affinity and Stability Data of **3** and Sequence Mutants[a]

sequence	wild type (**3**)	mutation	T_m (°C)	relative K_d
R	C8-G14	N_7		92
3	wild type		52	1
M1	C8C9T10T11	TTCC	52	342
M2	G14	A	42	4.5
M3	A13	G	49	11
M4	C12	T	46	4.5
M5	T11	C	57	820
M6	T10	C	51	750
M7	C9	T	49	1.9
M8	C8	T	53	4.5
M9	G4	removed	64	34
M10	C8-G14	mixed	55	480
M11	T6·A13	A·T	55	1.0
M12	C16	T		15
T7	C8-G14	$(dT)_7$		10

[a] Relative K_d values were measured using a gel shift assay and are normalized against **3**. The absolute apparent K_d value of the 11F8·**3** complex is 0.73 nM (per IgG binding site). **M10** contains the same base content in the selected region as **3** but in random order. **M12** contains the original constant region without the mutations introduced during the selection. **R** denotes the random pool using selection 2 primers (see Materials and Methods). Note that wild-type **3** and all mutants except **M9** are 55 bases long. Each affinity measurement was conducted at least three times, affording standard deviations ≤ 10% of the average apparent dissociation constant (K_d).

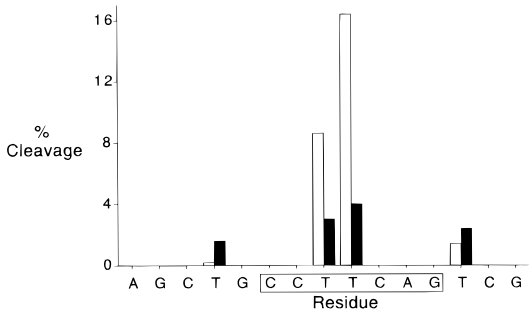


FIGURE 2: $KMnO_4$ mapping of sequence **3**. The white bars represent modification at 20 °C, and the black bars represent modification at 94 °C. The boxed residues highlight the selected region.

loop, separated from a bulge by three base pairs. Presumably, stem−loop sequences are selected in preference to those that are linear and necessarily more flexible because of the entropic benefit of preorganizing the single-stranded binding epitope (*19*).

Probing 11F8·Ligand Complexes. After the secondary structure of **3** was established, 11F8 complexes with this ligand and several sequence variants were probed by chemical footprinting to identify the residues important for recognition (Figure 3). The two loop thymines of **3** are protected from $KMnO_4$ modification when bound to 11F8, indicating that they interact with the protein. Residue T15 is also protected upon complex formation whereas T6 may be slightly more accessible to modification. These observations suggest that the stem duplex of **3** may undergo small structural rearrangements upon 11F8 binding. Reversing the loop to 5′-TTCC (**M1**) affords no protection of the loop thymines, suggesting that in addition to contacts with the bases either the loop conformation or location of the bases within the selected ligand is important for recognition. Because the NMR spectra of **1** do not provide evidence of structure in the loop as evidenced by hydrogen bonding, we favor the latter explanation.

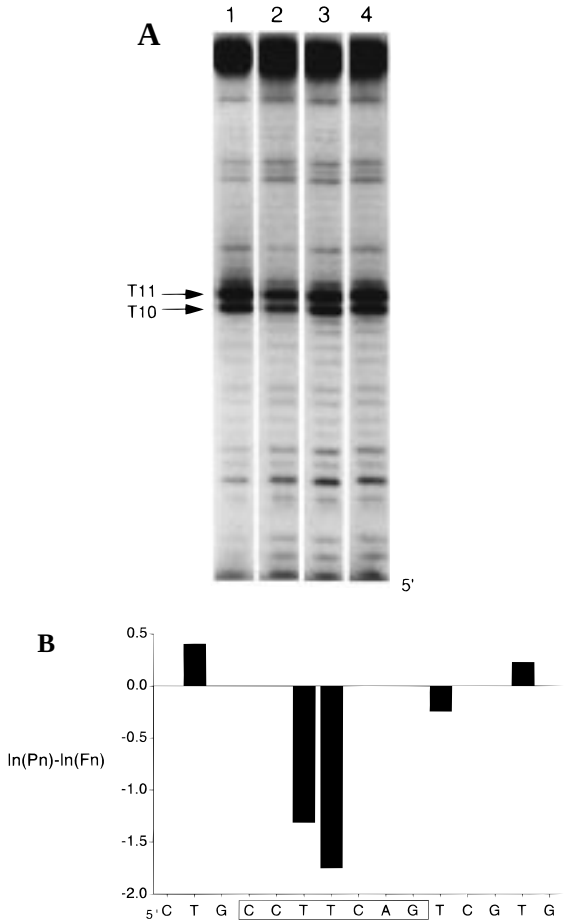


FIGURE 3: (A) $KMnO_4$ protection footprinting of **3** with normal mouse IgG (lane 1) and 11F8 (lane 2). For comparison, 9F11 and 15B10 are shown in lanes 3 and 4, respectively. (B) Quantification of the protection 11F8 affords **3** upon binding. They were analyzed as described by Rhodes (*14*). The boxed residues highlight the selected region. The difference in strand cleavage is represented by bars and is calculated by subtracting the logarithm of the probability of cleavage at each position in the free DNA (F_n) from the same position in the complex (P_n). A larger negative value of $\ln(P_n/F_n)$ indicates greater protection from $KMnO_4$ oxidation.

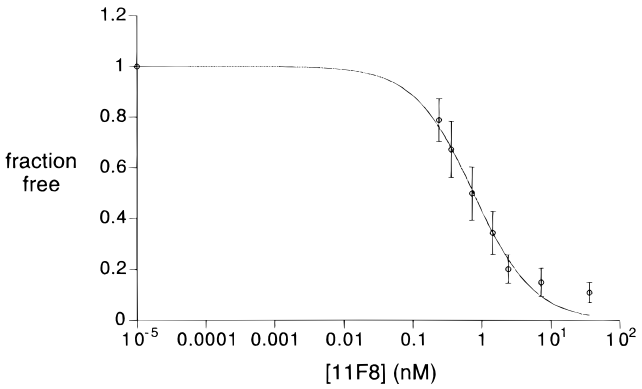


FIGURE 4: Representative binding isotherm of eight gel shifts of **3** and 11F8. The data were measured and analyzed as described in Materials and Methods. In the experiment shown, data were collected through 83% saturation. The error bars represent the standard deviation from eight separate replicates, and the solid line is the best fit to the data.

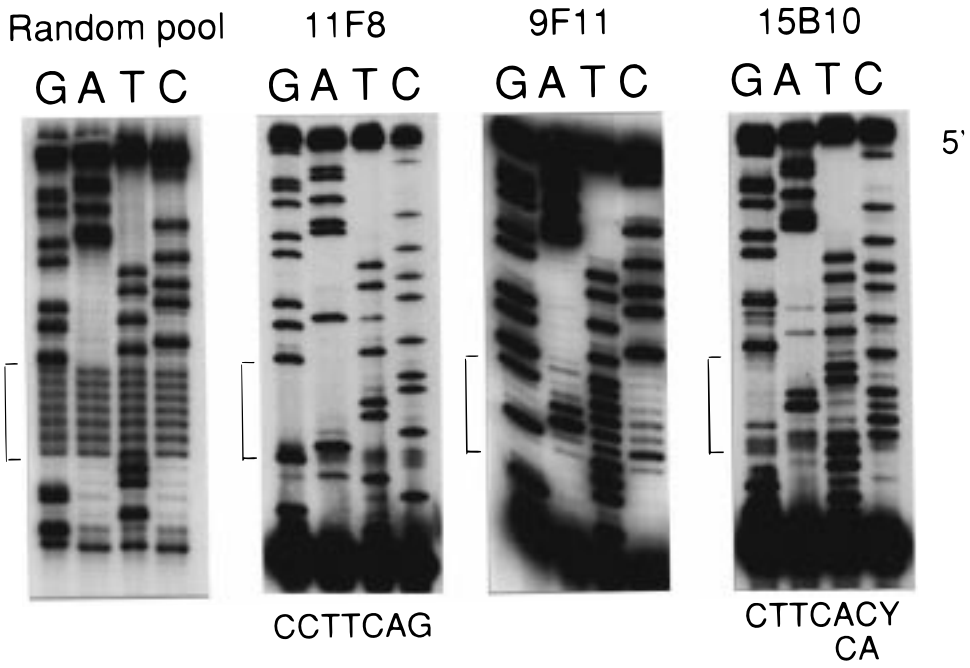
The affinity of 11F8 for **3** and 12 sequence mutants was measured using a gel shift assay (Table 1 and Figure 4). The apparent K_d value of 11F8 for **3** is 0.73 nM. This affinity is at least 100-fold higher than that observed for the initial pool and approaches a 10^3-fold improvement when compared with **M10** which possesses the same base sequence of **3** but in a random order. Mutant **M12** contains the selected sequence in the context of the wild-type flanking sequence, and it does not adopt a stem−loop secondary structure. Thus, the 15-fold decrease in affinity for this sequence provides an estimate of the energetic benefit of preorganization in **3** with respect to 11F8 binding. Although each of the point mutations tested decreases binding to some degree, mutations to T10 and T11 in the loop decrease affinity to the greatest extent. Changing the individual dC residues in the loop to dT does not improve binding, and 11F8 binds **3** nearly 10^3-fold tighter than it binds either **M1** or other hairpins containing pentathymidine loops.[3] Collectively, these experiments indicate that (i) the two loop thymines supply critical sequence-dependent recognition contacts for 11F8 and (ii) 11F8 is not just a "thymine-specific" binding protein. Both the sequence of **3** and context in which this sequence is presented to 11F8 are important for binding.

Deleting the bulge in **M9** results in a modest decrease in binding affinity. Because the selection data suggest that the bulge base itself does not interact with 11F8 (vide supra), it is possible that removing G4 eliminates functional groups, such as a proximal phosphate, that bind to the protein. To help address this point, ENU interference footprinting experiments were conducted to examine the interaction of 11F8 with the backbone of **3**. This study reveals that alkylation of about seven phosphates, one of which is adjacent to the bulge, interferes with binding (data not shown). To help distinguish whether ethylating the phosphates disrupts an electrostatic interaction or creates a steric hindrance to binding, the contribution of ion-pair formation to complex stability was assessed by measuring the K_d values as a function of $[Na^+]$ (*33*). Analysis of 11F8 binding at $[Na^+]$ ranging from 150 to 500 mM NaCl indicates that there is a net release of about one counterion upon binding (data not shown), which suggests that at most one phosphate forms a salt bridge in the 11F8·**3** complex. On the basis of the ENU footprinting data, it is possible that this phosphate is adjacent to the bulge. Hence, the decrease in affinity observed by removing the bulge could result from disrupting a phosphate contact with 11F8, possibly by altering the conformation of **3**.

Specificity of Other Anti-ssDNA mAbs. We also conducted selection experiments on mAbs 9F11 and 15B10 using the selection 2 construct (see Materials and Methods). These mAbs arose from the same B-cell precursor as 11F8, and as a group, they have identical light chains and differ by no more than seven somatic mutations in the hypervariable region of the heavy chain (*10*, *28*). Selections with 9F11 and 15B10 therefore have the potential to evaluate the effect of small variations in protein primary sequence on the fine specificity of anti-DNA·DNA interactions in vitro. Despite the similarity in amino acid sequence between the three mAbs, the DNA ligands isolated after nine rounds of selection/amplification with 9F11 and 15B10 are different from each other and are different from **3** (Figure 5). Cloning and sequencing revealed that in the final pool the 15B10 selection is dominated by 5′-CTTCACY (**5**, Y = C; **6**, Y = T). Two families of T-rich sequences are dominant in the

[3] Swanson, P. C., and Glick, G. D., unpublished observations.

A



B

INITIAL POOL		5' CTCGTGCTG**NNNNNNNN**TTGTCAC
3		5' CTCG<u>A</u>GCTG**CCTTCAG**T<u>C</u>GTCAC
5		5' CTCGT_CTG**CTTCACC**TTGTG GAC
6		5' CTCGT_CTG**CTTCACT**TTGTG GAC
7		5' CTCGTGCTG**TGTTTGT**TTGTGAC

FIGURE 5: (A) Sequencing gels of the final pools from selections with 11F8, 9F11, and 15B10. The data for all three mAbs was obtained using the selection 2 construct and primers (see Materials and Methods). Note that for 9F11 a single sequence cannot be read directly for sequencing the final pool as a whole. (B) Selected sequences for 11F8, 9F11, and 15B10. Sequence **3** is specific for 11F8, and **5** and **6** were selected for 15B10. Sequence **7** (and **6**) was selected for 9F11. The underlined bases represent mutations from the original constant region. Both bases in **3** changed from dT. The boxed bases are addition mutations, and a space denotes a deleted base. The selected regions are shown in bold.

9F11 experiments, of which 5′-CTTCACT (**6**) and 5′-TGTTTGT (**7**) are representative.

As observed with 11F8, mutations are present in the constant regions of **5** and **6** that presumably arose during PCR amplification of the selected DNA. However, the *absence* of mutations in the constant region of **7** suggests that those base changes/deletions introduced during PCR are *selected* by the mAbs. In other words, the base changes present in the final selected sequences are not random. Sequences **5**−**7** were evaluated for potential secondary structure using the chemical and spectroscopic methods described above. Although UV thermal denaturation studies show weak monophasic transitions, chemical mapping (*32*) does not indicate any base pairing. Taken together, we conclude that **5**−**7** do not adopt stable secondary structures; the conformational complementarity obtained through *binding-site* selection is unique to 11F8.

The three anti-ssDNA investigated here each bind their respective selected ligands at least 100-fold tighter than the initial random DNA pool, with 15B10 having the highest ratio. The degree to which a protein can discriminate between a "specific" site, those of similar composition, and random DNA is commonly considered the (thermodynamic) measure of specificity (*16*, *17*). Although a model describing sequence-specific recognition of ssDNA by proteins is not yet available, the ratio between specific and nonspecific binding affinities that is generally accepted to be the limit of sequence-specific recognition by dsDNA binding proteins range from 3 (*34*) to $\geq 10^2$ (*16*, *17*). Complexes between our mAbs and their corresponding selected ligands clearly meet this criterion of sequence specificity (Table 2).

In addition to their selected ligands, mAbs 11F8, 9F11, and 15B10 exhibit high affinity toward oligo(dT). Indeed, most proteins that bind ssDNA display a strong nonspecific preference for a particular homopolymer. Because discrimination between a homopolymer and sequence-specific ligand must represent a greater degree of specificity relative to random ssDNA (e.g., **R**; see Table 1), we propose that the preferred homopolymer for a given protein is the most appropriate comparison to assess sequence specificity.

Table 2: Apparent K_d Values Determined from Gel Shift Measurements[a]

sequence	9F11	15B10	11F8
R	27	500	67
T7	0.32	1.2	6.7
3	>40	>50	0.73
5	0.87	3.5	7.0
6	0.17	1.3	3.3
7	0.55	2.6	12

[a] The data are reported per IgG binding site (nM). Each affinity measurement was conducted at least three times, affording standard deviations ≤ 10% of the average apparent dissociation constant (K_d).

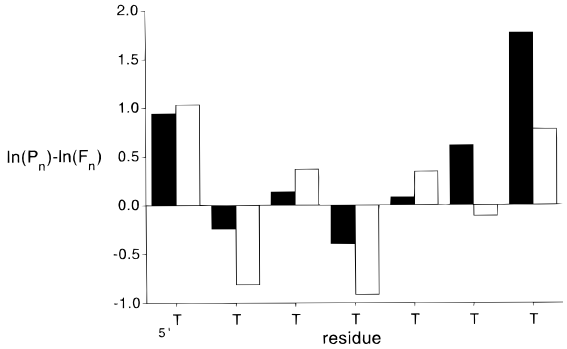


FIGURE 6: Quantification of representative $KMnO_4$ protection footprinting of 11F8 and 9F11 complexed with **T7**. The data for 15B10 are presented in ref *9* using $A_8-T_5-A_8$ as the ligand. Black bars = 11F8, and white bars = 9F11. There is no significant protection outside this region.

Through the use of a ligand that contains dT_7 flanked by the selection 2 constant region as a reference (**T7**), another set of specificity ratios were generated. In contrast to the ratios generated for 9F11 and 15B10 which are close to 1, 11F8 displays a 10-fold greater affinity for **3** than for **T7** (Table 2). On the basis of this comparison, we conclude that 11F8 possesses the greatest overall sequence specificity.

$KMnO_4$ footprinting experiments support our conclusions regarding sequence specificity based on affinity measurements. All three mAbs show a similar protection pattern when complexed to **T7**, which argues for a similar mode of (nonspecific) binding (Figure 6). However, footprinting an optimized ligand as in the 11F8·**3** or 9F11·**7** complexes each results in different protection patterns suggesting more specific binding interactions. For example, when bound to 11F8, T10 and T11 in **3** are protected from permanganate oxidation, which supports their importance in recognition of this ligand. By contrast, there is no pair of adjacent thymidine residues strongly protected in the 11F8·**T7** complex. 9F11 and 15B10 exhibit little or no protection of T10 and T11 in **3** (Figure 7a). Moreover, in the 9F11·**3** and 15B10·**3** complexes, increased reaction with permanganate is also observed for thymidine residues outside of the hairpin loop, particularly for 15B10, suggesting that the stem duplex opens upon binding. This induced fit is probably the source of the lower binding affinities observed for the two proteins. These data support our conclusions that preorganization of the recognition site is specific for 11F8, and that 11F8 possesses the greatest degree of sequence specificity.

In contrast to the interaction with **3**, the three mAbs display nearly identical protection patterns when complexed to **5** and **6** (Figure 7b). This observation suggests that although 9F11 and 15B10 bind to **5** and **6** with very high affinity as they

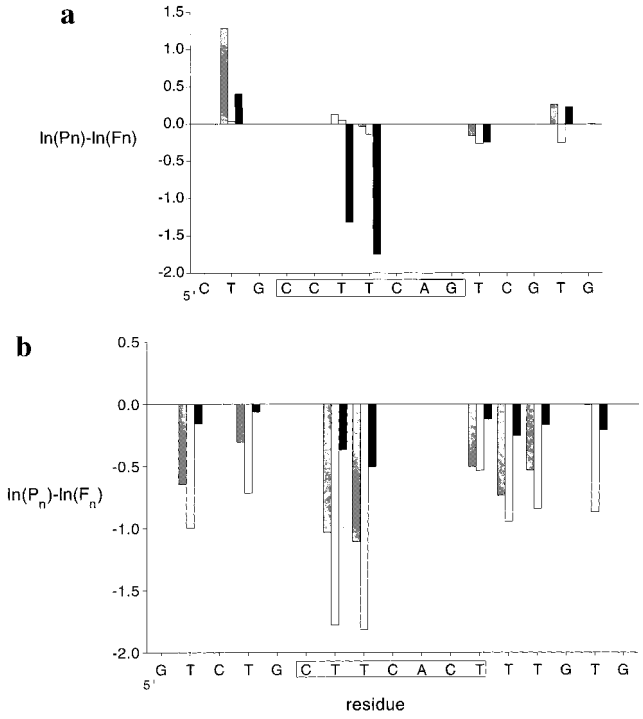


FIGURE 7: Quantification of the $KMnO_4$ protection for mAbs complexed with (a) **3** and (b) **6** (*14*). The boxed residues highlight the selected region. The gels from which these data are derived are shown in Figure 3. Using **5** in place of **6** affords the identical protection pattern. Black bars = 11F8, white bars = 9F11, and gray bars = 15B10. Boxed residues highlight the selected region. As seen in part b, nonspecific ligand **6** shows some protection outside the selected region and probably arises because the mAbs bind this sequence in multiple orientations.

do to **T7**, the interactions with these ligands are nonspecific. Without high-resolution structural data, the molecular basis for the specificity observed with the different mAbs interacting with the various ligands is unclear. Notwithstanding, the selection data indicate that considerations other than just absolute affinity must be considered when defining sequence-specific recognition by anti-ssDNA.

The selection experiments with the three mAbs clearly demonstrate that small changes in the amino acid sequence can give rise to large changes in specificity. Although it is not yet known if other anti-ssDNA display similar properties, analogous conclusions have been reached through site-directed mutagenesis experiments of anti-dsDNA (*35*). The somatic mutations in 11F8, 9F11, and 15B10 are not parallel, indicating that these mAbs have evolved along different pathways from a common progenitor B cell. Furthermore, because the germline genes that encode 11F8, 9F11, and 15B10 have not been isolated, we do not know which mAb has more somatic mutations than the others. However, the homology observed upon comparison of **3** with **5**−**7** may arise from affinity conferred by germline-encoded residues shared by the three mAbs. It is possible that with additional somatic mutations 9F11 and 15B10 may bind DNA with greater specificity.

A Novel Mode of Specific Protein·DNA Binding. Although 11F8 possesses considerable specificity for **3**, this mAb does bind other ssDNA molecules with appreciable affinity. This type of background recognition is a common feature of nucleic acid binding proteins in general, yet such "cross-reactivity" is not usually observed with IgG antibodies that

have undergone affinity maturation. MAb 11F8 also is unlike many well-characterized DNA binding proteins in other respects (*16*, *36*, *37*). In particular, electrostatic interactions often provide a significant component of stability for protein·DNA complexes, including those involving both specific and nonspecific recognition sequences (*36*, *38*). However, ion-pair formation in complexes of 11F8 and either **3** or oligo(dT) (*9*) is very similar. Collectively, therefore, it appears that specific recognition of **3** by 11F8 cannot be described solely using the usual paradigms invoked to explain other protein·DNA or antibody·antigen interactions.

Sequence-specific recognition of nucleic acids by proteins is often divided into two limiting mechanisms: direct and indirect readout site selection (*18*). These mechanisms involve matching complementary arrays of hydrogen bonds between DNA bases and a protein or recognition of a sequence-dependent conformation such as a subtle bend in duplex DNA, respectively (*18*, *38*). Although recognition of dsDNA is generally dominated by the direct readout mechanism, small variations in local DNA conformation can dramatically influence protein binding by helping to present functional groups in a manner that is complementary to the protein (*19*, *38*). In contrast to DNA, RNA commonly folds into a diverse range of secondary structures. As a result of this rich structural diversity, RNA binding proteins rely heavily on the indirect readout mechanism. Not unlike the interaction between 11F8 and **3**, RNA binding proteins often target the single-stranded regions of preorganized secondary structures to achieve a high degree of specificity (*18*, *39*, *40*). Therefore, to describe specific recognition of DNA by anti-ssDNA, aspects of both direct and indirect readout must be invoked.

CONCLUSIONS

We have identified preferred recognition sequences for several anti-ssDNA mAbs. To the best of our knowledge, these experiments represent the first application of in vitro selection technology to explore the sequence specificity of ssDNA binding proteins. By using the concepts associated with protein·DNA, protein·RNA, and antibody·antigen recognition, we have started to describe aspects of sequence-specific recognition of ssDNA by anti-ssDNA that may extend to other proteins that recognize ssDNA. Indeed, ssDNA binding proteins have been identified recently that have regulatory functions in eukaryotic cellular metabolism, and these proteins are proposed to interact with ssDNA in a sequence-specific fashion (*41*−*43*). Our results may provide a starting point for the further understanding and characterization of the interaction of these proteins with ssDNA. Structural and thermodynamic studies to further elucidate the molecular basis for the observed specificity of 11F8 are currently in progress.

SUPPORTING INFORMATION AVAILABLE

Description of NMR spectroscopy and resonance assignments (3 pages). Ordering information given on any current masthead page.

REFERENCES

1. Wilson, I. A., and Stanfield, R. L. (1993) *Curr. Opin. Struct. Biol. 3*, 113−118.
2. Roitt, I., Brostoff, J., and Male, D. (1989) in *Immunology*, 2nd ed., Gower Medical Publishing, London.
3. Tan, E. M. (1989) *Adv. Immunol. 44*, 93−151.
4. Koffler, D., Schur, P. H., and Kunkel, H. G. (1967) *J. Exp. Med. 126*, 607−624.
5. Waer, M. (1990) *Clin. Rheumatol. 9, Suppl. 1*, 111−115.
6. Isenberg, D. A., Ehrenstein, M. R., Longhurst, C., and Kalsi, J. K. (1994) *Arthritis Rheum. 37*, 169−180.
7. Ben-Chetrit, E., Eliat, D., and Ben-Sasson, S. A. (1988) *Immunology 65*, 479−485.
8. Pisetsky, D. S. (1992) *Rheum. Dis. Clin. North Am. 18*, 437−454.
9. Swanson, P. C., Ackroyd, P. C., and Glick, G. D. (1996) *Biochemistry 35*, 1624−1633.
10. Blatt, N. B., Bill, R. M., and Glick, G. D. (1998) *Hybridoma 17*, 33−39.
11. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) in *Molecular Cloning. A Laboratory Manual*, 2nd ed., Vol. 3, Cold Spring Harbor Laboratory Press, Plainview, NY.
12. Stevens, S. Y., Swanson, P. C., and Glick, G. D. (1994) *J. Immunol. Methods 177*, 185−190.
13. Swanson, P. C., Cooper, B. C., and Glick, G. D. (1994) *J. Immunol. 152*, 2601−2612.
14. Rhodes, D. (1988) in *Protein Function: A Practical Approach* (Creighton, T. E., Ed.) IRL Press, Oxford,.
15. Ross, S. A., and Burrows, C. J. (1996) *Nucleic Acids Res. 24*, 5062−5063.
16. von Hippel, P. H., and Berg, O. G. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 1608−1612.
17. von Hippel, P. H. (1994) *Science 263*, 769−770.
18. Draper, D. E. (1995) *Annu. Rev. Biochem. 64*, 593−620.
19. Eaton, B. E., Gold, L., and Zichi, D. A. (1995) *Chem. Biol. 2*, 633−638.
20. Szostak, J. W. (1992) *Trends Biomed. Sci. 17*, 89−93.
21. Ouellette, M. M., and Wright, W. E. (1995) *Curr. Opin. Struct. Biol. 6*, 65−72.
22. Blackwell, T. K., and Weintraub, H. (1990) *Science 250*, 1104−1110.
23. Thiesen, H.-J., and Bach, C. (1990) *Nucleic Acids Res. 18*, 3203−3209.
24. Tao, X., and Murphy, J. R. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 9646−9650.
25. Henderson, B. R., Menotti, E., Bonnard, C., and Kühn, L. C. (1994) *J. Biol. Chem. 269*, 17481−17489.
26. Andrews, B. S., Eisenberg, R. A., Theofilopoulos, A. N., Izui, S., Wilson, C. B., McConahey, P. J., Murphy, E. D., Roths, J. B., and Dixon, F. J. (1978) *J. Exp. Med. 148*, 1198−1215.
27. Tuerk, C., and Gold, L. (1990) *Science 249*, 505−510.
28. Swanson, P. C., Yung, R., Eagan, M., Norris, J., Blatt, N. B., Johnson, K., Richardson, B. C., and Glick, G. D. (1996) *J. Clin. Invest. 97*, 1748−1760.
29. Aboul-ela, F., Koh, D., and Tinoco, I., Jr. (1985) *Nucleic Acids Res. 13*, 4811−4824.
30. LeBlanc, D. A., and Morden, K. M. (1991) *Biochemistry 30*, 4042−4047.
31. Wang, Y.-H., and Griffith, J. (1991) *Biochemistry 30*, 1358−1363.
32. McCarthy, J. G. (1989) *Nucleic Acids Res. 17*, 7541.
33. Record, M. T., Jr., Lohman, T. M., and de Haseth, P. (1976) *J. Mol. Biol. 107*, 145−158.
34. Kurumizaka, H., Kanke, F., Matsumoto, U., and Shindo, H. (1992) *Arch. Biochem. Biophys. 295*, 297−301.
35. Katz, J. B., Limpansithikul, W., and Diamond, B. (1994) *J. Exp. Med. 180*, 925−932.
36. Kowalczykowski, S. C., Bear, D. G., and von Hippel, P. H. (1981) in *The Enzymes* (Boyer, P. D., Ed.) Vol. XIV, Academic Press, Inc., London.
37. Janin, J. (1995) *Proteins 21*, 30−39.
38. Neidle, S. (1994) in *DNA Structure and Recognition*, IRL Press, Oxford.

39. Nagai, K., and Mattaj, I. W. (1994) in *Fronteirs in Molecular Biology* (Hames, B. D., and Glover, D. M., Eds.) 1st ed., Oxford University Press, Oxford, U.K.
40. Puglisi, J. D., Tan, R., Calnan, B. J., Frankel, A. D., and Williamson, J. R. (1992) *Science 257*, 76−80.
41. Kneale, G. G. (1992) *Curr. Opin. Struct. Biol. 2*, 124−130.
42. Stevens, S. Y. (1997) Ph.D. Thesis, University of Michigan, Ann Arbor, MI.
43. Usheva, A., and Shenk, T. (1996) *Proc. Natl. Acad. Sci. U.S.A. 93*, 13571−13576.

BI981899O