Recognition Properties of a Sequence-Specific DNA Binding Antibody†

James F. LeBlanc, Kathryn E. McLane,‡ Paul W. H. I. Parren, Dennis R. Burton, and Peter Ghazal*

Departments of Immunology and Molecular Biology, The Scripps Research Institute, 10666 North Torrey Pines Road, La Jolla, California 92037

Received December 15, 1997; Revised Manuscript Received February 18, 1998

ABSTRACT: A sequence-specific DNA-binding antibody was previously generated by incorporating a 17 amino acid α-helix from the DNA-binding domain of the transcription factor TFEB into the HCDR3 site of a recombinant human Fab fragment. The recombinant DNA-binding antibody, called Fab-E box, binds the TFEB recognition sequence CACGTG (an E box site) with a 5−10-fold lower affinity than TFEB. Here, we have determined the precise kinetics of interaction of Fab-E box with DNA and show that the lower affinity of Fab-E box relative to TFEB for E box DNA is due to a higher dissociation rate. DNase I protection assays show Fab-E box physically interacts with one half-site of the E box. Additional DNA target sites of Fab-E box were identified by DNase I protection assays. A compilation of these binding sites indicates that the recognition elements for Fab-E box binding include a half-site of the E box, CAW, with an 8 bp consensus sequence identified as YNYYCAWW. Thus, the DNA determinants for Fab-E box recognition extend beyond one-half site of the E box sequence, with preferences for pyrimidines and A+T-rich sequences in the 5′ and 3′ outer regions of the binding site, respectively. Apparent dissociation constants of Fab-E box for a subset of these target DNA sequences are 5−10-fold greater than the DNA-binding affinity of the antibody with the E box site. Therefore, these results identify important DNA specificity determinants for high-affinity binding by Fab-E box.

Antibodies are a convenient target for protein engineering to produce molecules with designed binding specificities. Immunoglobulins and Fab fragments are very stable proteins whose overall structures can tolerate a wide variety of amino acid sequences in their antigen-binding domains. Antibodies that bind DNA with high affinity have been identified as an important component of the autoimmune syndrome of systemic lupus erythematosus, SLE (*1*−*3*). These naturally occurring molecules are not targeted to a specific DNA motif and appear to bind to a wide variety of DNA sequences with little specificity. Autoantibodies that bind double-stranded DNA (dsDNA) have been proposed to primarily recognize the helical pattern of the phosphate groups with ionic interactions through basic amino acids in their antigen-binding sites (*4*−*7*). In addition, a natural antibody repertoire to dsDNA for generating molecules with altered binding specificities is not practical as dsDNA is not highly immunogenic (*8*).

We have used phage display, in an initial attempt to overcome this problem, to select DNA-binding Fab fragments in vitro from a library which was constructed by random mutagenesis of the heavy chain CDR3 (*9*). This method was successful in isolating Fab fragments that bound repetitive poly(dG)·poly(dC) and poly(dG-dC)·poly(dG-dC) sequences, but not Fabs which recognize specific sequences. To generate a sequence-specific molecule, we designed a new class of DNA-binding antibody by cloning a 17 amino acid DNA recognition α-helix from a well-characterized eukaryotic transcription factor, TFEB, into the HCDR3[1] domain of a human Fab fragment. This protein, called Fab-E box, recognizes the E box site CACGTG via interactions in the major groove of dsDNA (*10*). Equilibrium competition studies indicated that Fab-E box bound the E box site with an affinity that is approximately 1 order of magnitude lower than the parental transcription factor, and that the sequence recognition was less restricted than that of TFEB (*10*). Thus, while these studies demonstrated the ability of an engineered antibody to recognize a specific DNA sequence for the first time, the question of whether antibodies can bind DNA with an affinity and specificity comparable to known DNA-binding proteins remains open.

Here, we present evidence that the DNA-binding antibody Fab-E box recognizes DNA with both high affinity and specificity. Precise kinetic and equilibrium measurements of the antibody−DNA interactions show that the lower affinity (5−10-fold) of the antibody for DNA containing the E box sequence compared to the affinity of TFEB for E box DNA is due to the differences in the dissociation rates of the two DNA−protein complexes. DNase I protection experiments were used to characterize the specificity of the Fab-E box−DNA interactions and to more precisely define

† This work was supported by DARPA, and P.G. is a Scholar of the Leukemia Society of America. Publication no. 11461-1MM of The Scripps Research Institute.

* Corresponding author. E-mail address: ghazal@scripps.edu.

‡ Present address: Genentech Inc., South San Francisco, CA.

[1] Abbreviations: HCDR3, heavy chain complementary determining region; bp, base pair(s); nt, nucleotide(s); DTT, dithiothreitol; HEPES, 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid; EDTA, (ethylenedinitrilo)tetraacetic acid; EMSA, electrophoretic mobility shift assays; Y, the pyrimidines C or T; W, the DNA bases A or T; IPTG, isopropyl β-D-thiogalactoside; AdMLP, adenovirus major late promoter; SDS, sodium dodecyl sulfate; Tris, tris(hydroxymethyl)aminomethane; AMP, adenosine monophosphate; EDC, *N*-ethyl-N′-[(dimethylamino)propyl]-carbodiimide hydrochloride; NHS, *N*-hydroxysuccinimide.

S0006-2960(97)03079-1 CCC: $15.00
Published on Web 04/04/1998

the boundaries of the DNA sequences recognized by the antibody. These protection experiments with E box-containing DNA showed that the antibody recognizes a half-site of the E box dyad symmetry. Intriguingly, these experiments further revealed that the Fab fragment can also selectively bind other related DNA sequences. Affinity measurements of Fab-E box binding to a subset of these variant binding sites demonstrated that the antibody can specifically recognize DNA sequences with an affinity comparable to naturally occurring DNA-binding proteins. These high-affinity binding sites for the Fab fragment appear to consist of one or several E box half-sites, with the consensus site identified as YNYYCAWW. The results of this study strongly support the feasibility of generating sequence-specific DNA-binding antibodies.

MATERIALS AND METHODS

Protein Purification. Fab-E box and TFEBΔ265 were expressed and purified as described except that 1 mM IPTG and 6 μM cyclic AMP were used to induce transformed *E. coli* for only 3 h before harvesting during Fab production (*10*). TFEBΔ265 will be referred to as TFEB in the text.

Electrophoretic Mobility Shift Assays (*EMSA*). TFEB and Fab-E box were incubated with the indicated concentrations of competitor DNA at room temperature in volumes of 15−20 μL in binding buffer (25 mM HEPES, pH 7.5, 5 mM $MgCl_2$, 50 mM KCl, 5% glycerol, 5 mM DTT, 0.1% Nonidet P-40, with 0.1−0.5 ng of [α-^{32}P] end-labeled target DNA). Binding reactions with TFEB included 0.1 mg mL^{-1} poly(dG-dC)·poly(dG-dC) as nonspecific competitor. Reactions were incubated for 30−45 min before being loaded onto 4% polyacrylamide gels [39:1 acrylamide:bis(acrylamide) ratio] cast and electrophoresed in 25 mM Tris base and 190 mM glycine (pH 8.3). Gels were prerun for 30 min at 150 V (∼15 mA) prior to loading samples. Electrophoresis was performed at 10−15 mA for 30−45 min at room temperature. Gels were dried, autoradiographed, and/or exposed to a Phosphoimager (Molecular Dynamics) screen for data acquisition. ImageQuant software (Molecular Dynamics) was used for data collection while the Kaleidagraph package (Abelbeck software) was used for graphical analysis and determination of binding constants.

DNase I Footprinting. AdMLP-containing plasmid DNA for footprinting of the E box was a kind gift of Diane Hawley, University of Oregon. TFEB and Fab-E box were incubated with 0.1−0.5 ng of [α-^{32}P] end-labeled DNA (10 000−50 000 cpm) in a volume of 20−25 μL in binding buffer for 30−45 min at room temperature. DNA digestion was performed for 30 s with RQ1 DNase I (Promega, 0.05−0.08 unit) diluted into binding buffer with 2 mM $CaCl_2$. DNase I digestion was stopped by the addition of 100 μL of 0.5% SDS, 1 M ammonium acetate, and 0.15 mg mL^{-1} yeast RNA. Stopped reactions were vortexed, phenol/chloroform-extracted, ethanol-precipitated, washed with 75% ethanol, and dried under vacuum. Dried pellets were resuspended in 95% deionized formamide plus running dyes, heated at 100 °C for 2−3 min, and loaded onto a sequencing gel (6−10%) cast in 1× TBE (90 mM Tris, 90 mM boric acid, 2.5 mM EDTA, pH 8.3) with 8 M urea. Gels were dried onto 3 mm filter paper and autoradiographed overnight at −80 °C with a DuPont Cronex Lightning Plus intensifying screen.

Surface Plasmon Resonance (*BIAcore*) *Assays*. Kinetic constants for binding of Fab-E box and TFEB to biotinylated dsDNA were determined using surface plasmon resonance in a BIAcore instrument from Pharmacia. Streptavidin (200 μg/mL in 10 mM sodium acetate, pH 4.5) was coupled to the (carboxymethyl)dextran matrix of a CM5 sensor chip (Biosensor) using EDC and NHS amide coupling (*11*). After coupling, unreacted NHS ester groups were inactivated with 1 M ethanolamine hydrochloride (pH 8.0). Typically, 5000 BIAcore response units (RU) of streptavidin were immobilized.

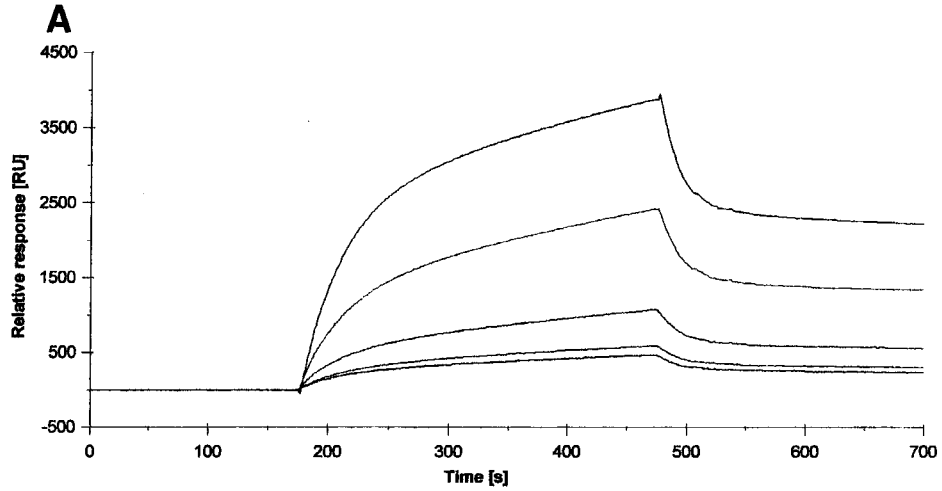
Synthetic oligonucleotides corresponding to the E box (AdMLP sequences −45 to −69), MLP1 and MLP2 (*10*), were obtained commercially (Operon). Single-stranded 5′-biotinylated MLP1 and MLP2 oligonucleotides were annealed to the corresponding nonbiotinylated complement, and 40 μL of the mixture of both 5′-biotinylated dsDNA probes (700 pg/mL) was injected for capture onto the streptavidin-coupled sensor chip.

Binding analyses were performed in binding buffer in the presence of 0.005% surfactant P20 at a constant flow rate of 5 μL/min. The association rate (k_{on}) was determined by measuring the rate of binding of the Fab-E box and TFEB to the sensor chip surface at different protein concentrations. The dissociation rate (k_{off}) was determined by increasing the flow rate to 50 μL/min after the association phase. The k_{on} and k_{off} values were calculated using BIAcore kinetics evaluation software. The k_{on} was obtained from k_{obs}, the slope of the plot d(RU)/dt versus RU, at different protein concentrations. The plot of ln (RU1/RUt) versus time (t) was used to obtain k_{off}. Values for K_D (k_{off}/k_{on}) were determined using the average k_{off} and k_{on} values obtained from four independent measurements.

RESULTS

Measurement of DNA-Binding Kinetics of TFEB and Fab-E Box. Fab-E box was shown from our previous work to bind to a dsDNA 28-mer containing the E box sequence with ∼10-fold lower affinity than the parental transcription factor, TFEB (*10*). In the present study, we have used surface plasmon resonance to compare the binding kinetics of TFEB and Fab-E box to the immobilized dsDNA AdMLP probe (described under Materials and Methods). The binding kinetics of TFEB and Fab-E box were measured at various protein concentrations (Figure 1). Binding of TFEB and Fab-E box to streptavidin alone was negligible (data not shown). Figure 1A is the BIAcore sensorgram for binding of five different concentrations of TFEB (25−400 nM). Figure 1B is the sensorgram for the binding of Fab-E box at four different antibody concentrations (200−800 nM).

The kinetic analyses from these experiments indicated that the principal difference in the binding of TFEB and Fab-E box to E box DNA is the stability of the DNA−protein complex formed. The association rate determined for both DNA-binding proteins with the E box at 25 °C was very similar (k_{on}= 2.8 × 10^4 M^{-1} s^{-1} for TFEB and 2.5 × 10^4 M^{-1} s^{-1} for Fab-E box), but the complexes formed by the two proteins had about a 3-fold difference in their off rates (k_{off} = 1.4 × 10^{-3} s^{-1} for the TFEB−DNA complex and 3.7 × 10^{-3} s^{-1} for the Fab−DNA complex, Figure 1). A



C

	k_{on} (s^{-1} M^{-1})	k_{off} (s^{-1})	K_D (nM)
TFEB	2.8 (± 0.7) x 10^4	1.4 (± 0.3) x 10^{-3}	49 ± 8
Fab-E box	2.5 (± 0.7) x 10^4	3.7 (± 0.1) x 10^{-3}	160 ± 55

FIGURE 1: Determination of rate constants for E box DNA binding by TFEB and Fab-E box by surface plasmon resonance (BIAcore) analysis. (A) Sensorgram overlay plot of TFEB association with and dissociation from E box DNA over time. The lines represent the binding of 25, 50, 100, 200, and 400 nM TFEB. (B) Sensorgram overlay plot of Fab-E box association with and dissociation from E box DNA over time. The lines represent the binding of 200, 400, 600, and 800 nM Fab-E box. (C) Table of DNA binding kinetic constants determined from BIAcore experiments ($n = 4$).

kinetic measurement of the overall binding constant is determined by the ratio of the on and off rates ($K_D = k_{on}/k_{off}$), and so the difference between the K_D for TFEB and Fab-E box, 49 nM to 160 nM, is due to the instability of the Fab−DNA complex relative to that formed by TFEB with the E box.

Equilibrium Determinations of Antibody−DNA Affinities. Previously, rough estimates by EMSA competition assays of the affinities of TFEB and Fab-E box for the E box sequence were 2 and 20 nM, respectively (*10*). These values, although showing a similar trend as the BIAcore determinations, differed in absolute terms, and so more rigorous EMSA equilibrium measurements were performed. These band-shift assays were different from those used in the initial studies of Fab-E box as the conditions for running native polyacrylamide gels were changed from low ionic strength Tris−borate (TBE) buffers to high ionic strength Tris−glycine buffers (see Materials and Methods). These new electrophoresis conditions improved the amounts of Fab-complexed DNA detected in the gel 3−5-fold (data not shown). Titration studies using these different buffer conditions for the EMSA experiments agreed with the previous estimate of the binding affinity difference between TFEB and the antibody. Using the double-reciprocal method of determining the overall equilibrium dissociation constant, TFEB was measured to bind oligomers containing E box DNA with a K_D of 23 ± 4 nM while the K_D of Fab-E box interactions with the same sequence was 200 ± 50 nM (Figure 2). These results agree with the BIAcore measurements, and thus all subsequent affinity measurements were performed with this modified EMSA protocol.

Examination of Antibody Binding Specificity. Previous studies with Fab-E box were restricted to analyzing DNA binding to short oligonucleotides by EMSA. Accordingly, we next investigated antibody−DNA interaction by performing DNase I footprinting assays on a larger fragment of DNA

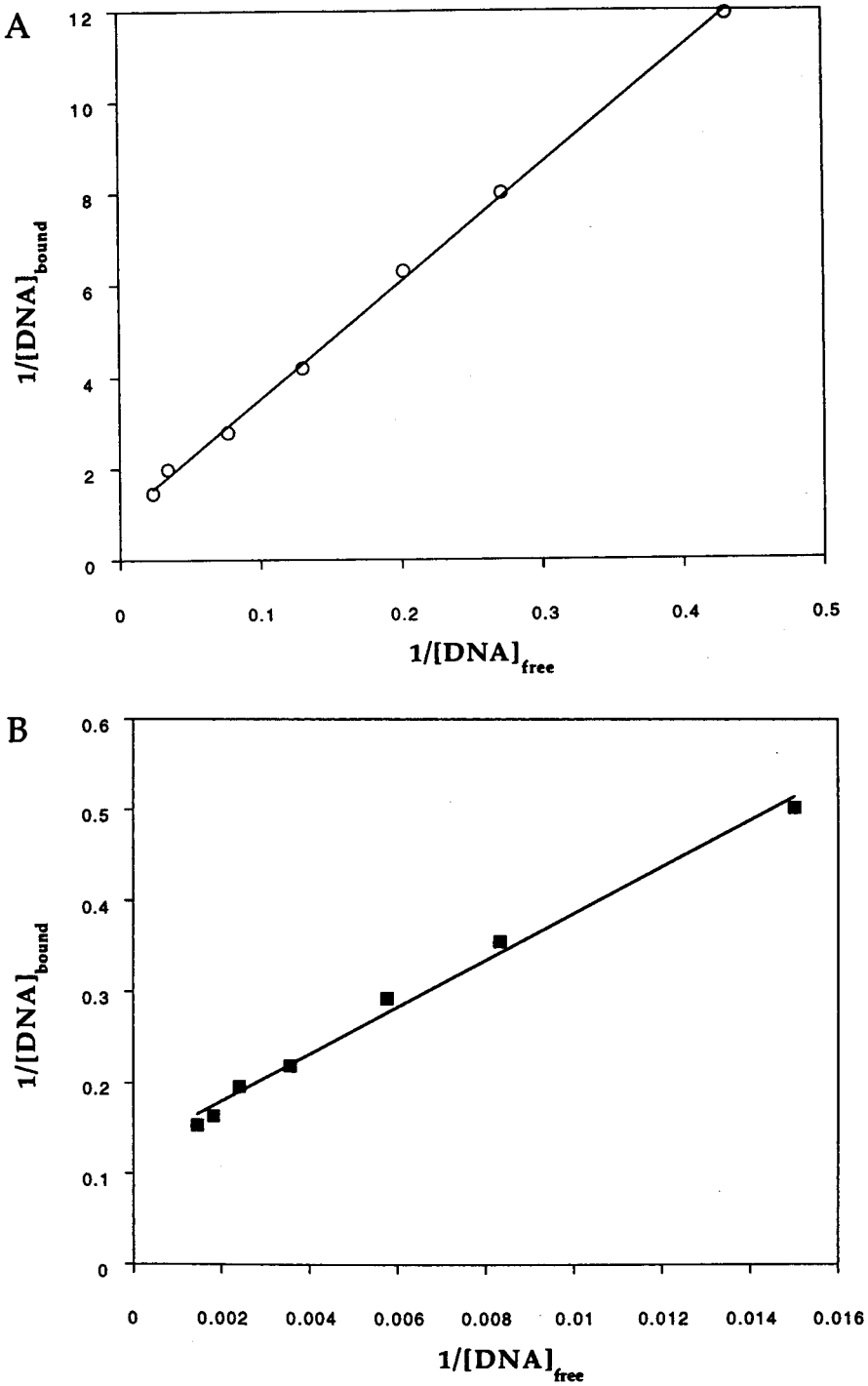


FIGURE 2: Determination of binding constants using DNA competition titrations with electrophoretic mobility shift analysis (EMSA) of TFEB and Fab-E box binding to the E box DNA element from the Adenovirus Major Late Promoter (AdMLP sequences −48 to −68). (A) Results of a DNA-binding titration experiment with TFEB to the E box element as measured by EMSA. The double-reciprocal plots of three separate binding experiments give a K_D of 23 ± 4 nM. (B) Results of a DNA-binding titration experiment with Fab-E box to the E box element as measured by EMSA. The double-reciprocal plots of five separate binding experiments give a K_D of 200 ± 50 nM.

to determine the sequences protected by Fab-E box binding. A [^{32}P] end-labeled fragment from the AdMLP was used for some of the initial footprinting experiments as the short oligomers used for the EMSA experiments contain the E box sequence centered around the central CACGTG motif (AdMLP sequences −55 to −60). The AdMLP E box sequence has been shown to bind a variety of transcription factors from the basic helix−loop−helix (bHLH) family of DNA-binding proteins including USF, c-Myc, Max, TFE3, and TFEB (*12*−*19*). As expected, TFEB protects a single large region from DNase I digestion centered over the E box from −49 to −66 (Figure 3A), consistent with previous protection experiments (*16*). In contrast, Fab-E box developed a split or bipartite protection pattern on both DNA strands (Figure 3B,C). The E box motif itself is protected by the antibody, but it appears that Fab-E box preferentially binds the upstream half-site of the E box. These results therefore further demonstrate the specificity of Fab-E box binding and define the DNA-binding site for the E box sequence as half of the dyad symmetry.

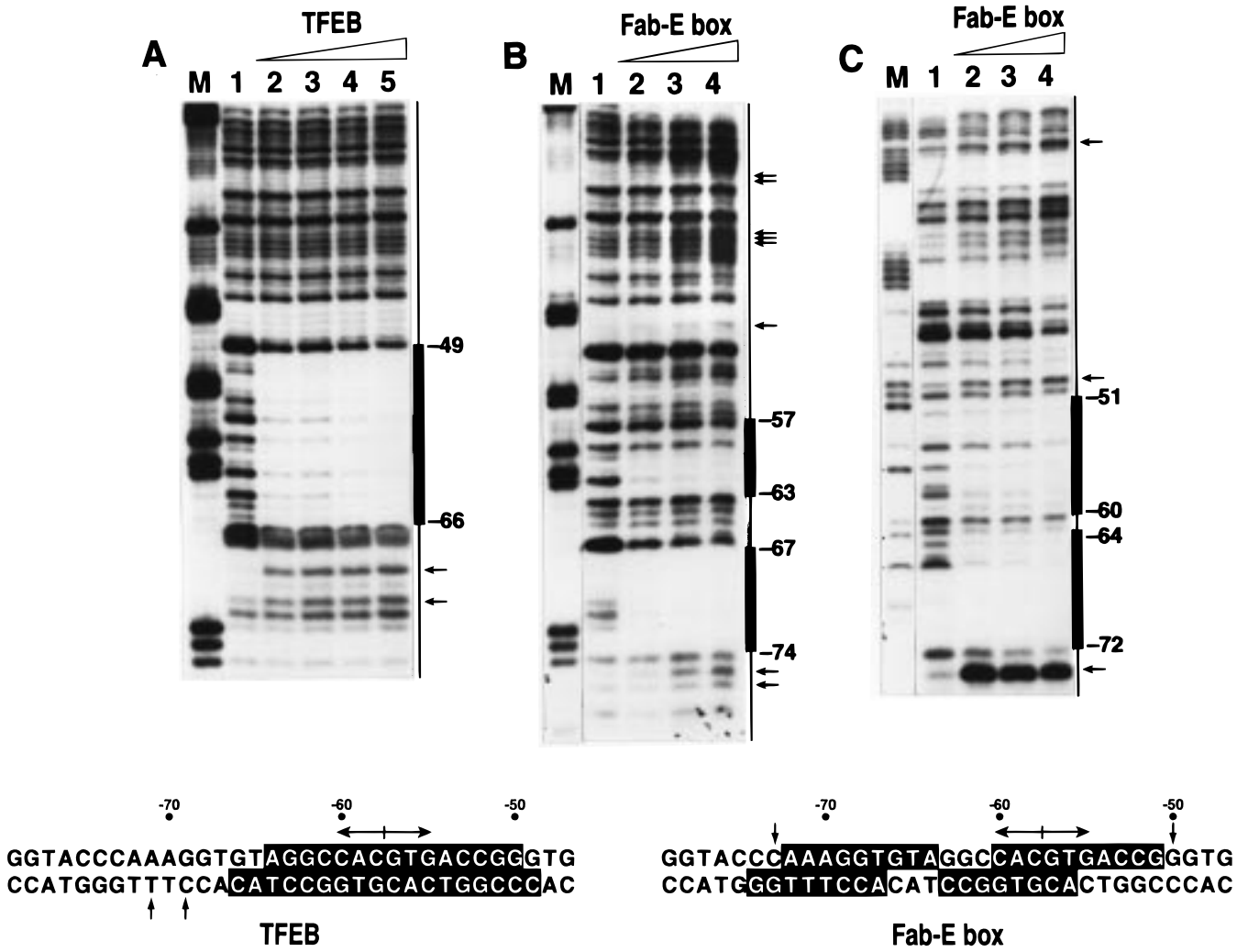


FIGURE 3: DNase I footprinting analysis of TFEB and Fab-E box binding to the E box element of the Adenovirus Major Late Promoter (AdMLP). Panel A shows the protection pattern of TFEB on the bottom strand of the AdMLP. Lane M indicates a guanosine chemical sequencing reaction of the end-labeled fragment for a marker lane. Lanes 1−5 contain reactions with 0, 0.25, 0.5, 1, and 2 μM TFEB, respectively. Panels B and C show DNase I protection assays by Fab-E box of the bottom and top strands, respectively, of the AdMLP. Lanes 1−4 contain reactions with 0, 150, 300, and 600 nM Fab-E box. The boundaries of sequences protected are indicated by black rectangles while arrows show sites of DNase I hypersensitivity upon protein binding. A schematic representation of the footprinting data is shown below.

Unexpectedly, sequences located immediately 5′ of the E box were protected by Fab-E box while TFEB makes little, if any, contact with these upstream sequences. (Compare lane 5 in Figure 3A to lane 4 in Figure 3B.) The binding of this upstream AdMLP site by Fab-E box protects a larger region than the E box and induces a very pronounced DNase I hypersensitive site on the top strand near −73 (compare lanes 1 and 2, Figure 3C). Protection of this upstream site also appeared to occur at lower Fab-E box concentrations from the protein titrations with the AdMLP shown in panels B and C of Figure 3. These results suggest that Fab-E box makes more extensive contacts with this upstream site than with the E box itself.

These upstream AdMLP sequences were not originally included in the oligomers used in the EMSA experiments that measured the equilibrium dissociation constant of the Fab fragment for the E box (K_D = 200 ± 50 nM, Figure 2B). These sequences upstream of the E box were introduced into an oligomer containing the complete antibody-binding site termed the 'Fab box' (AdMLP sequences −50 to −75). EMSA competition binding assays with this new oligomer determined that the equilibrium dissociation constant for Fab-E box to this extended site is K_D = 53 ± 9 nM. These experiments thus revealed a 4-fold increase in binding affinity compared to the E box alone by the inclusion of these upstream sequences. The fact that TFEB does not recognize these AdMLP sequences suggests that other domains of Fab-E box may be involved in DNA−protein interactions. For instance, additional protein contacts outside of the TFEB recognition α-helix on the heavy chain could assist Fab-E box in binding non-E box DNA.

It then appears that the Fab fragment has an altered DNA-binding specificity from TFEB and there may be other sequences that are even more tightly bound by the antibody than those found upstream of the AdMLP E box. To investigate this possibility, DNase I protection assays were carried out with other DNA fragments to search for DNA-binding sites for Fab-E box. The results from one of these experiments is shown in Figure 4A, where Fab-E box was found to protect sequences downstream of the AdMLP. Lanes 1 and 2 of Figure 4A show the results of DNase I digestion (on the bottom strand) of these downstream DNA sequences

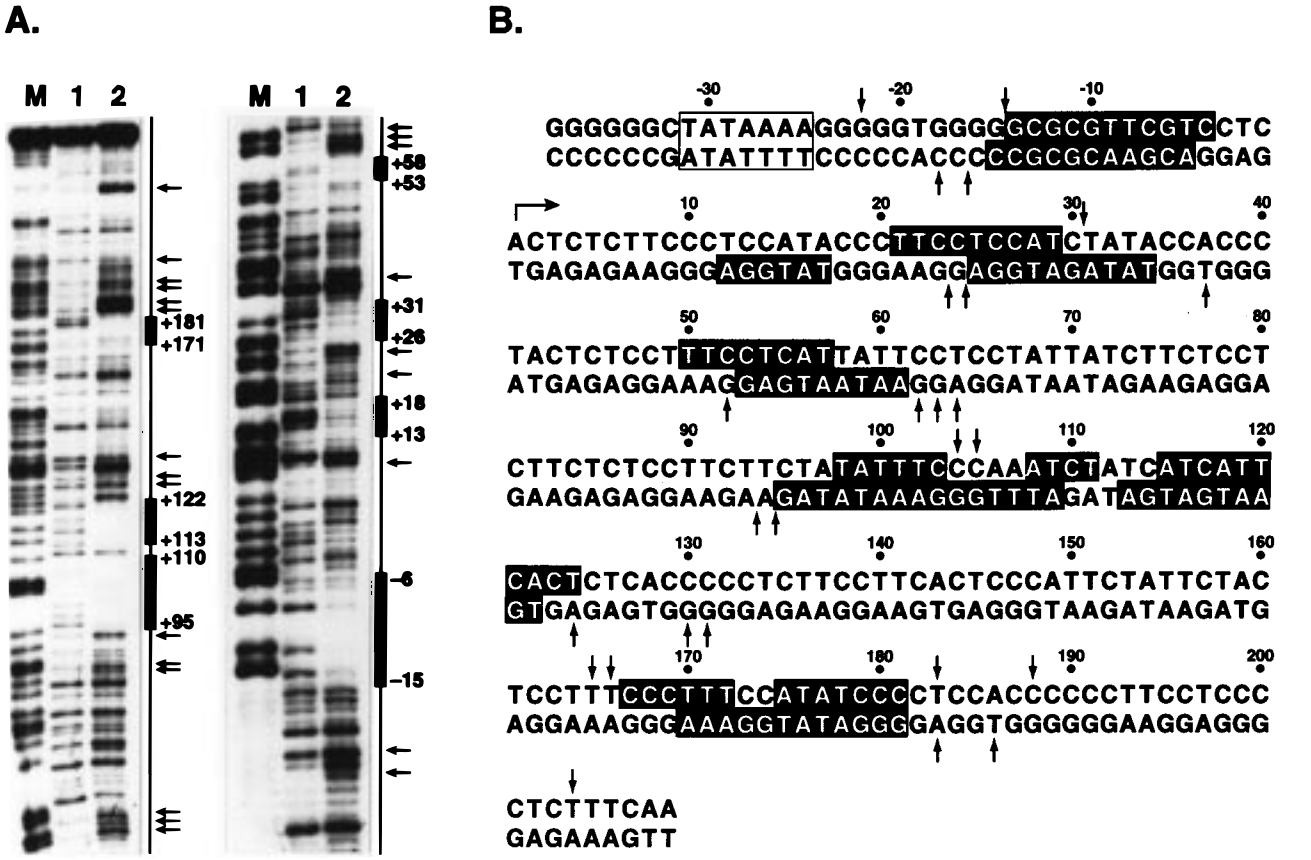
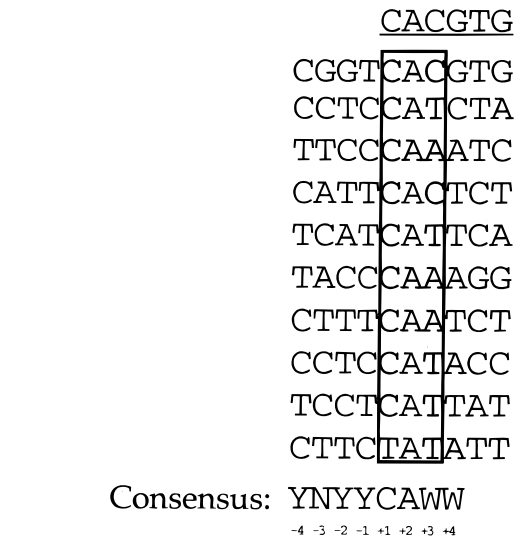


FIGURE 4: DNase I footprinting analysis of Fab-E box binding to DNA sequences downstream of the AdMLP. Panel A shows the protection of the downstream boundary of the promoter (sequences to +7) by Fab-E box DNA binding. Lane M is a marker lane made by guanosine-specific chemical cleavage of a ^{32}P end-labeled fragment; lanes 1 and 2 have the results of DNase I cleavage of the fragment after incubation in the absence and presence of 300 nM Fab-E box, respectively. The schematic depiction of Fab protection and resulting hypersensitive sites are given in panel B.

in the absence and presence of Fab-E box. Figure 4B gives a schematic representation of the protection and hypersensitive patterns due to Fab binding on both strands downstream of the AdMLP TATA box. The binding data represented in Figure 4B were collected from a series of experiments to best give an accurate picture of the size and location of the protected sites. These protected regions generally appear to be between 6 and 10 bases in size with hypersensitive sites on 1 or both of the protected boundaries. The exceptions to the general protected size limit of 6−10 base pairs are 2 longer sites, 1 of about 15 bases centered around +105 and another of about 12−15 bases centered near +175 (Figure 4B). Examination of these longer sites reveals that the protection pattern is broken up on one of the strands; the site around +105 leaves about four bases unprotected on the top strand, and the site near +175 also leaves two bases available for DNase I cleavage between two protected regions. These observations suggest that these sequences, like the 'Fab box' site from −50 to −74 of the AdMLP, may actually consist of adjacent antibody-binding sites. The results of these and other protection assays with different sequences were collected and analyzed to determine a possible consensus sequence for Fab-E box recognition (see Table 1). This collection of antibody-binding sites reveals a consensus binding sequence of YNYYCAWW, indicating a conservation of an E box half-site, CAW, along with a preference for two A/T bases at the 3′ edge of the E box half-site. However, some sequences have been shown not to be recognized by Fab-E box even though they have a high degree of sequence homology with the putative consensus.

Table 1: DNA Sequences Bound by Fab-E Box As Demonstrated by DNase I Protection Assays[a]



[a] The E box sequence bound by TFEB is shown underlined at the top while a potential consensus sequence recognized by the antibody is shown at the bottom. The sources of the sequences shown include the E box of the AdMLP and the regions protected from DNase I cleavage shown in Figure 4.

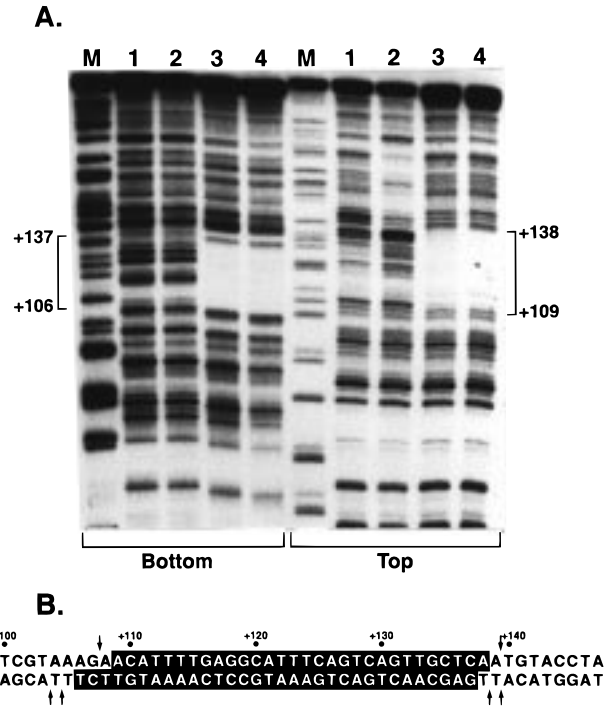


FIGURE 5: Identification and characterization of a large high-affinity site bound by Fab-E box. Panel A shows the DNase I protection patterns of the bottom and top strands in the 5′ portion of the chloramphenicol acetyltransferase (CAT) gene by the binding of Fab-E box and TFEB. Lane M is a guanosine-specific sequencing reaction of the labeled fragment, while lane 1 is a control lane showing DNase I cleavage in the absence of any protein. Lanes 2, 3, and 4 are the protection patterns due to the presence of 500 nM TFEB and 20 nM and 40 nM Fab-E box, respectively. The numbers assigned to the bases are given respective to the start of transcription from a eukaryotic promoter upstream of the CAT gene. The sequence of this site, termed the 'YCA box', along with a depiction of Fab protection and induced hypersensitive sites of both strands is given in panel B. The determination of the equilibrium binding constant, K_D, was done by EMSA as described in Figure 2. The results of four separate binding experiments with a 44 bp fragment containing the YCA box give a K_D of 17 ± 7 nM.

Similarly, sequences bearing little resemblance to YNYYCAWW are bound by the antibody (such as GCGCGTTCGTC found between −14 and −4 of the AdMLP in Figure 4B). These cases demonstrate that flanking sequences, DNA structure, and the juxtaposition of binding sites may also play important roles in DNA recognition by Fab-E box.

Other binding experiments examining DNase I protection of nonpromoter DNA led to the discovery of a high-affinity Fab-E box binding site in the 5′ portion of the chloramphenicol acetyltransferase (CAT) gene (Figure 5). Figure 5A shows the DNase I pattern on the bottom and top strands of the CAT gene after incubation with TFEB (lane 2) and Fab-E box (lanes 3 and 4). A schematic of the protection pattern of this binding site is shown in Figure 5B. The first striking observation of this experiment is that TFEB appears to have little, if any, interaction with this sequence while the antibody with the same DNA recognition α-helix strongly protects this site at a much lower protein concentration. Another conspicuous feature of Fab-E box binding to this site is the large size of the protected region, from +109 to +138 on the top strand and from +106 to +137 on the bottom strand, making it the largest site protected from DNase I digestion by the antibody that has been detected (compare the size of the protected sites from Figures 4 and 5).

This binding site has been named the 'YCA box' due to the presence of a number of (C/T)CA trinucleotide motifs that are found in the Fab-E box consensus sequence (Table 1). Oligomers containing the YCA box were synthesized, annealed, and end-labeled with ^{32}P for DNA titration EMSA experiments to determine the equilibrium dissociation constant of Fab-E box to this sequence. Double-reciprocal analyses of four separate binding experiments were used to calculate a K_D of 17 ± 7 nM. Identical results were obtained by Langmuir isotherm analysis of the binding data. This high-affinity site appears to be collection of E box half-sites bound by a number of Fab molecules. This suggests that the increased affinity could be due to the more extensive protein−DNA contacts along the length of this site as well as possible protein−protein interactions between adjacent Fab fragments bound to the YCA box.

DISCUSSION

We have analyzed the DNA-binding activity of an engineered DNA-binding antibody, Fab-E box, by DNase I footprinting, EMSA, and surface plasmon resonance binding (BIAcore) assays. The results of this study demonstrate that Fab-E box recognizes about 8 bp of DNA with a consensus sequence of YNYYCAWW. Our experiments also show that Fab-E box is capable of binding certain DNA sequences with affinities comparable to naturally occurring DNA-binding proteins.

DNase I protection experiments showed that Fab-E box recognizes one half-site of the E box sequence. This observation invites comparison to the recognition properties of the bHLH family of DNA-binding proteins. The recognition of the canonical CANNTG E box sequence involves base-specific interactions of each bHLH monomer primarily with the conserved two bases at each edge of the site (*14*, *20*, *21*). Other studies have also indicated that some E box binding proteins have sequence preferences outside of the conserved 6 bp motif (*22*−*26*). ADD1, another member of the bHLH family, and TFEB have a strong binding preference for a T immediately 5′ of the CA in the E box (*22*, *27*). Therefore, Fab-E box appears to recognize DNA much like a bHLH protein monomer would be expected; it has strong preferences for a central CA sequence with a pyrimidine immediately upstream. The binding specificity of Fab-E box is not limited to just this YCA motif, as there are strong preferences for pyrimidines at two other positions (−2 and −4) and an adenine or thymine at positions +3 and +4 (Table 1).

These DNA recognition properties are similar to a protein related to the basic leucine zipper (bZIP) family that binds DNA as a monomer, Skn-1 from *C. elegans* (*28*). Skn-1, which lacks a dimerization domain (as does Fab-E box), binds a half-site from the bZIP family consensus sequence and prefers an AT-rich region immediately upstream. Skn-1 has a greater stringency for binding DNA than does Fab-E box, with two sets of 5 bp consensus sites identified, one centered around the sequence GTCAT and another around ATCAT. Taken together, it appears that Skn-1 binding requires four specific bases with a strong preference for an upstream YYR sequence. The looser specificity of Fab-E box compared to a monomer binding protein such as Skn-1

is not surprising when it is considered that the antibody possesses only a 17 amino acid segment taken from the basic domain of TFEB. In comparison, Skn-1 has an 85 residue C-terminal DNA-binding domain as well as a short N-terminal domain presumed to make contacts in the minor groove of DNA (*28*).

The next step in the evolution of these novel DNA-binding factors will take advantage of the dimerization of Fab fragments. The fact that the majority of DNA-binding proteins are dimers attests to the fact that specificity can be greatly increased when the binding energy difference between specific and nonspecific sites is high. Dimers typically bind two DNA half-sites separated by a distinct spacing that greatly limits the number of nonspecific sites bound. The presumed ability of the Fab fragment to bind DNA as a monomer limits the free energy price paid by binding only a single E box half-site; therefore, the affinity difference between binding sites is smaller for Fab-E box than the parental TFEB dimer. A recent report demonstrated an increase of approximately 10-fold in the specificity of helix−loop−helix peptides binding E box-containing DNA when the proteins were dimerized by a disulfide bridge (*29*). The specificity increase expected with a linked pair of DNA-binding domains in the same molecule should also limit the number of 'nonspecific' DNA sites bound by Fab-E box. Therefore, this study provides the important first step in demonstrating that antibody molecules can be used to generate DNA-binding proteins with high affinity and specificity for a selected target sequence.

ACKNOWLEDGMENT

We acknowledge the valuable work of Raiza Bastidas and Roman Rozenshteyn for the production of Fab-E box and Carl Haarstad for his generous assistance in FPLC purification of TFEBΔ265. We also thank Alan MacLaughlin for his help in identifying the Fab-E box consensus sequence.

REFERENCES

1. Deutscher, S. L., and Keene, J. D. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 3299−3303.
2. Stollar, B. D., and Schwartz, R. S. (1985) *J. Clin. Invest. 75*, 321.
3. Tan, E. M. (1989) *Adv. Immunol. 44*, 93−152.
4. Braun, R. P., and Lee, J. S. (1987) *J. Immunol. 139*, 161.
5. Braun, R. P., Woodsworth, M. L., and Lee, J. S. (1986) *Mol. Immunol. 23*, 685−691.
6. Herron, J. N., He, X.-H., Ballard, D. W., Blier, P. R., Pace, P. E., Bothwell, A. L. M., Voss, E. W., Jr., and Edmundson, A. B. (1991) *Proteins*: *Struct., Funct., Genet. 11*, 159−175.
7. Mol, C. D., Muir, A. K. S., Lee, J. S., and Anderson, W. F. (1994) *J. Biol. Chem. 269*, 3605−3610.
8. Stollar, B. D. (1986) *CRC Crit. Rev. Biochem. 20*, 1−36.
9. Barbas, S. M., Ghazal, P., Barbas, C. F., III, and Burton, D. R. (1994) *J. Am. Chem. Soc. 116*, 2161−2162.
10. McLane, K. E., Burton, D. R., and Ghazal, P. (1995) *Proc. Natl. Acad. Sci. U.S.A. 92*, 5214−5218.
11. Jönsson, U., Fägerstam, L., Ivarsson, B., Johnsson, B., Karlsson, R., Lundh, K., Löfas, S., Persson, B., Roos, H., Sjölander, Stenberg, E., Stahlberg, R., Urbaniczky, C., Ostlin, H., and Malmqvist, M. (1991) *BioTechniques 11*, 620−627.
12. Carthew, R. W., Chodosh, L. A., and Sharp, P. A. (1985) *Cell 43*, 439−448.
13. Chodosh, L. A., Carthew, R. W., and Sharp, P. A. (1986) *Mol. Cell. Biol. 6*, 4723−4733.
14. Ellenberger, T., Fass, D., Arnaud, M., and Harrison, S. C. (1994) *Genes Dev. 8*, 970−980.
15. Fisher, D. E., Carr, C. S., Parent, L. A., and Sharp, P. A. (1991) *Genes Dev. 5*, 2342−2352.
16. Fisher, D. E., Parent, L. A., and Sharp, P. A. (1993) *Cell 72*, 467−476.
17. Sawadogo, M., and Roeder, R. G. (1985) *Cell 43*, 165−175.
18. Sawadogo, M. (1988) *J. Biol. Chem. 263*, 11994−12001.
19. Sawadogo, M., Van Dyke, M. W., Gregor, P. D., and Roeder, R. G. (1988) *J. Biol. Chem. 263*, 11985−11993.
20. Ferré-D′amaré, A. R., Prendergast, G. C., Ziff, E. B., and Burley, S. K. (1993) *Nature 363*, 38−45.
21. Ferré-D′amaré, A. R., Pognonec, P., Roeder, R. G., and Burley, S. K. (1994) *EMBO J. 13*, 180−189.
22. Halazonetis, T. D., and Kandil, A. N. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 6162−6166.
23. Prendergast, G. C., and Ziff, E. B. (1991) *Science 251*, 186−189.
24. Fisher, F., and Goding, C. R. (1992) *EMBO J. 11*, 4103−4109.
25. Blackwell, T. K., Huang, J., Ma, A., Kretzner, L., Alt, F. W., Eisenmann, R. N., and Weintraub, H. (1993) *Mol. Cell. Biol. 13*, 5216−5224.
26. Bonven, B. J., Nielsen, A. L., Norby, P. L., Pedersen, F. S., and Jorgensen, P. (1995) *J. Mol. Biol 249*, 564−575.
27. Kim, J. B., Spotts, G. D., Halvorsen, Y. D., Shih, H. M., Ellenberger, T., Towle, H. C., and Spiegelman, B. M. (1995) *Mol. Cell. Biol. 15*, 2582−2588.
28. Blackwell, T. K., Bowerman, B., Priess, J. R., and Weintraub, H. (1994) *Science 266*, 621−628.
29. Kunne, A. G. E., and Allemann, R. K. (1997) *Biochemistry 36*, 1085−1091.

BI9730792