2760–2769 Nucleic Acids Research, 1999, Vol. 27, No. 13 © 1999 Oxford University Press

SELEX and missing phosphate contact analyses reveal flexibility within the AP-2α protein:DNA binding complex

Neeman Mohibullah, Amy Donner, Joseph A. Ippolito[1] and Trevor Williams*

Department of Molecular Cellular and Developmental Biology and [1]Department of Molecular Biophysics and Biochemistry, Yale University, 266 Whitney Avenue, New Haven, CT 06511, USA

Received February 11, 1999; Revised and Accepted May 10, 1999

ABSTRACT

The AP-2 family of transcription factors are defined by the presence of a novel DNA binding domain, termed a 'basic helix–span–helix' motif. The AP-2 genes regulate important aspects of vertebrate embryogenesis and have also been linked to the control of cell proliferation and tumorigenesis, but the cellular targets that the AP-2 proteins control are largely undefined. In particular, since only a limited number of sequences have previously been utilized to define the nature of the AP-2 binding site, the range of DNA sequences recognized by the AP-2 proteins remains unknown. We have therefore utilized a SELEX analysis to identify multiple new AP-2α binding sites. Moreover, we have devised a novel missing phosphate and nucleotide competition analysis to characterize the residues in the binding site required for AP-2α protein:DNA contact. These studies suggest that the AP-2α protein:DNA complex is flexible and indicate that AP-2α can bind three related sequence motifs: GCC N3 GGC, GCC N4 GGC and GCC N3/4 GGG. The availability of these more refined consensus sequences should assist in the identification of target genes for this critical transcription factor.

INTRODUCTION

The AP-2 family of transcription factors, AP-2α, AP-2β and AP-2γ (also termed AP-2.2), regulate important aspects of vertebrate development. These three genes are expressed in tissues undergoing complex morphogenetic changes during vertebrate embryogenesis, principally in the developing neural crest, neural tube, kidney, eye, facial prominences and limb buds (1–4). Knockout mouse studies have been performed so far on the AP-2α and AP-2β genes and indicate that both are required for normal embryogenesis. Mice lacking the AP-2β gene die shortly after birth, due to inappropriate morphogenesis of the kidneys (5). The disruption of the AP-2α gene leads to defects in the formation of the neural tube, face, eyes, body wall and limbs and results in perinatal lethality (6–9). The developmental programs affected in AP-2α knockout mice show a significant overlap with those caused by teratogenic levels of retinoic acid (10). Interestingly, the AP-2α and AP-2γ genes are also retinoic acid responsive, which further suggests an important link between AP-2 expression and the mechanism of action of this teratogen (11–14).

One key mechanism by which the AP-2 proteins may function in development is by regulating cell proliferation. The AP-2α gene can inhibit cell cycle progression (15) and aberrant expression of the AP-2 genes has been observed in human breast cancer and melanoma (1,16,17). The influence of AP-2 proteins on cell fate determination may be linked to their activation of genes involved in cell growth, cell communication and cell movement, including p21$^{waf1/cip1}$, c-erbB-2, E-cadherin and matrix metalloproteinases (15,18–22). The AP-2α protein can also interact with the *trans*-acting proteins c-myc, E1A, SV40 large T antigen and retinoblastoma (20,22–24). The interaction of AP-2α with such regulators may also provide an important mechanism for integrating cell proliferation, apoptosis and differentiation.

The AP-2 proteins define a distinct class of transcription factor characterized by a large, ~200 amino acid, C-terminal DNA binding domain (1,25,26). This DNA binding domain is bipartite and consists of a basic region necessary for DNA contact and a large dimerization motif of novel structure, termed a helix–span–helix. Consistent with the high degree of sequence identity found within the DNA binding domain, all three AP-2 proteins can bind to the same sequence motif, previously represented by the consensus sequence 5'-GCCNNNGGC-3' (25). However, this consensus sequence was inferred from a limited number of AP-2 binding sites and, more recently, several potential AP-2 sites have been proposed which are quite different in their sequence organization (27). Therefore, to understand more fully the nature of the AP-2 binding site we have isolated multiple new AP-2 recognition sequences using a PCR-based SELEX approach. In addition, we have characterized critical components of the AP-2 binding site using a variety of competition studies, including a novel missing phosphate analysis. Taken together, these studies reveal significant modifications from the original AP-2 consensus sequence and provide a more comprehensive picture of the AP-2α DNA binding site. This analysis has important implications both for identifying possible AP-2 target genes and also for understanding how AP-2 binds DNA at the molecular level.

*To whom correspondence should be addressed. Tel: +1 203 432 3230; Fax: +1 203 432 5690; Email: twill@biomed.med.yale.edu

MATERIALS AND METHODS

Oligonucleotides, PCR and sequencing

All oligonucleotides were manufactured by the Keck Facility at the Yale University School of Medicine. The binding site selection protocol was performed using a 58mer oligonucleotide, select1, 5'-GGG AGT CAA GCT TCC CAC GAG-(N)16-GGT CAC CTC GAA TTC GCA CCG-3', which contained a random core of 16 nt. The primer select3, 5'-CGG TGC GAA TTC GAG GTG ACC-3', was used for the initial synthesis of double-stranded DNA. Subsequently, for PCR amplification, the select3 primer was coupled with the primer select2, 5'-GGG AGT CAA GCT TCC CAC GAG-3'. The PCR reactions were performed with *Taq* DNA polymerase using conditions recommended by the manufacturer (Boehringer Mannheim). Amplifications were performed in 16 cycles of 95°C for 40 s, 62°C for 1 min, 72°C for 1 min, followed by an incubation at 72°C for 9 min. In the double-stranded DNA produced by these protocols, the core nucleotides are flanked by *Hin*dIII and *Eco*RI recognition sites.

A PCR-based approach was also utilized to identify bacterial clones containing the appropriate SELEX plasmids. A 1.5 ml aliquot of a bacterial culture was harvested by centrifugation and resuspended in 60 µl of single colony lysis buffer (SCLB; 10 mM Tris–HCl pH 7.5, 1 mM EDTA and 50 µg/ml proteinase K). The mixture was then incubated sequentially at 55°C for 15 min, 80°C for 15 min and 4°C for 1 min. The supernatant was harvested following centrifugation in a microfuge and 9 µl used for the PCR reaction. The T7 primer and a modified T3 primer, revT3, 5'-CAG CTA TGA CCA TGA TTA CGC C-3', were utilized for PCR. Amplifications were performed in 30 cycles of 95°C for 40 s, 48°C for 1 min, 72°C for 1 min, followed by 9 min at 72°C. PCR products were purified using Qiaquick columns (Qiagen) and sequenced with the revT3 primer by the Keck Facility at the Yale University School of Medicine.

AP-2α protein preparation

The bacterial expression vector for protein preparation, VFH ΔN165, was derived from the plasmid AP-2 linker which contains the human AP-2α cDNA (25). The 5'-end of the VFH ΔN165 construct was formed by ligating an *Nde*I linker (5'-CCC ATA TGG G-3') to the *Sma*I site of AP-2α located at nt 557, while the 3'-end was derived from the *Bam*HI site of the AP-2 linker. This fragment was then inserted into the vector pET3a (28) between the *Nde*I and *Bam*HI sites to create VFH ΔN165. The ΔN165 AP-2α protein was overexpressed in *Escherichia coli* BL21 cells containing the plasmid VFH ΔN165. The induced cells were lysed in buffer containing 25 mM HEPES (pH 7.5), 800 mM KCl, 12.5 mM $MgCl_2$, 20% glycerol, 5 mM DTT and lysozyme (at 1 mg/ml). Following lysis and removal of cell debris by ultracentrifugation, AP-2α was precipitated from the cell lysate supernatant by addition of $(NH_4)_2SO_4$ to 55% saturation. The AP-2α pellet was resolubilized in a solution containing 25 mM HEPES (pH 7.5), 0.02% NaN_3 and 10% glycerol and passed over a Q-Sepharose anion exchange column. The resulting flow-through, which was rich in AP-2α, was then loaded onto an SP-Sepharose cation exchange column. AP-2α was eluted with 600 mM KCl and was further purified on a Mono-S column attached to a Pharmacia FPLC system. Resulting fractions containing AP-2α protein were pooled and dialyzed against 10 mM HEPES (pH 7.5), 20 mM NaCl, 5 mM DTT, 0.02% NaN_3, 10% glycerol. The final AP-2α protein fraction was >95% pure as judged by Coomassie staining of SDS–PAGE gels (data not shown).

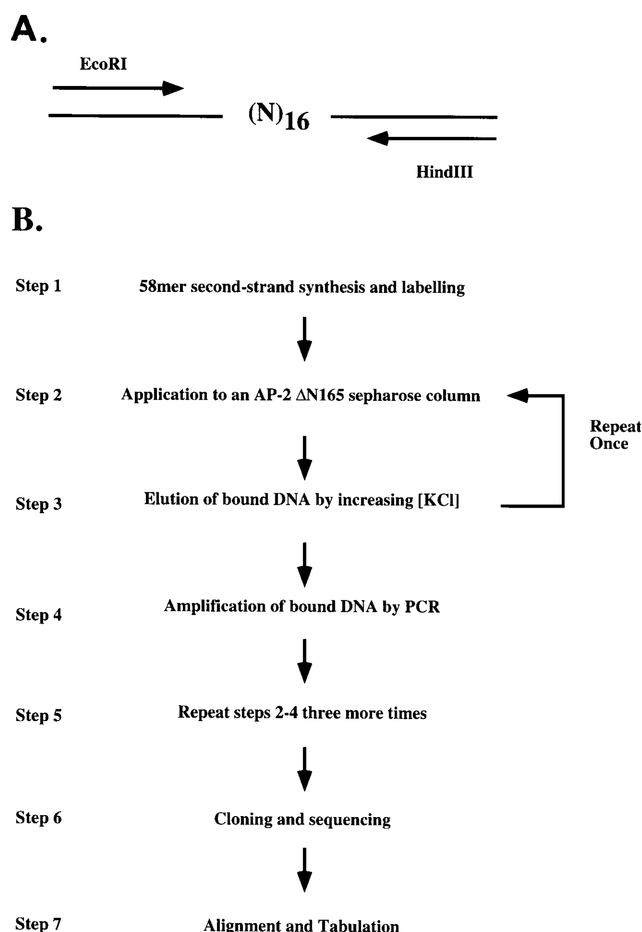


Figure 1. Strategy for SELEX analysis of the AP-2α binding site. (**A**) Schematic representation of the 58mer oligonucleotide used for binding site selection, showing the random core of 16 nt, the primer binding sites and the positions of the *Eco*RI and *Hin*dIII restriction sites. (**B**) Scheme for binding site selection.

Affinity purification of oligonucleotides containing AP-2α binding sites

The strategy chosen for the SELEX analysis of the AP-2α binding site (Fig. 1) was a modification of the procedure employed for the study of Ubx DNA recognition (29). The AP-2α ΔN165 protein was coupled to CNBr-activated Sepharose FF (Pharmacia) in 50 mM HEPES (pH 7.8), 0.5 M NaCl for 10 h at 4°C. Subsequently, the remaining active groups in the resin were inactivated by treatment with 0.1 M Tris–HCl (pH 7.9) as recommended by the manufacturer. Coupling efficiency was >90% and the final concentration was 3.6 mg ΔN165 AP-2 protein per ml resin. The resin was stored in 0.5 M NaCl, 0.1 M Tris (pH 7.9) and washed with buffer T.1 [50 mM Tris–HCl (pH 7.9), 100 mM KCl, 10 µg/ml gelatin, 10 µg/ml poly(dI·dC)/(dC·dI), 1 mM DTT] prior to oligonucleotide

Table 1. Percentage of input DNA eluted at a given KCl molar concentration (shown in parentheses)

Round	Flow-through and wash	Elution
1	99% (0.375)	1% (1.0)
2	95% (0.5)	3% (1.0)
3	90% (0.55)	10% (1.0)
4	80% (0.6)	4% (1.0)

addition. The double-stranded oligonucleotide pool for the first pass was made by annealing a 3-fold molar excess of select3 to select1 and then extending with Klenow DNA polymerase I (New England Biolabs) in the presence of cold TTP, dGTP, dCTP (500 µM each), dATP (100 µM) and [α-^{32}P]dATP. Subsequently, the labeled DNA was gel purified on an 8% acrylamide gel and ~10 µg applied to a 1 ml column at 4°C in 1 ml of buffer T.1. The column was washed successively with buffer T containing 0.1 (T.1), 0.25 (T.25), 0.375 (T.375) and 1 M (T1) KCl. The 1 M KCl fraction (1.24% yield) was diluted to T.25 and reapplied to the column. The resin was washed with buffer T.375 and the tightly bound fraction eluted with T1. The total yield from the first and second rounds of passage over the column was 1% (Table 1). An aliquot of the 1 M fraction was diluted 10-fold with water and 1 µl (~5 pg) utilized for PCR. Following gel purification, the oligonucleotides were again applied to the column, washed and eluted as outlined in Table 1. The PCR and column purifications were repeated twice more. After the final round of PCR, the oligonucleotide pool was digested with restriction endonucleases *Eco*RI and *Hin*dIII and cloned into the corresponding sites of pBS II SK– (Stratagene).

Electrophoretic mobility shift assay (EMSA)

EMSAs (25) and competition studies (30) were performed as described previously. These experiments utilized a 5'-end radiolabeled AP-2 probe derived from the human metallothionen IIa gene distal BLE (see below and Table 7) and either wild-type AP-2α protein or ΔN165 AP-2α protein, which were both derived from a rabbit reticulocyte lysate translation system (Promega). For missing phosphate contact analysis, the oligonucleotides utilized to study single-stranded breaks were based on the complementary 24mer sequences PhosX, 5'-CAA GAT CGC CTC GGG CTT AGA CAC-3', and PhosY, 5'- GTG TCT AAG CCC GAG GCG ATC TTG-3'. Competitor double-stranded oligonucleotides containing single-stranded breaks were prepared by annealing equimolar quantities of PhosX and the pair of derivatives corresponding to the PhosY strand. Annealing was performed in 100 mM NaCl, 10 mM Tris–HCl (pH 7.9), 1 mM EDTA by heating to 95°C for 1 min and then slowly cooling to room temperature. The hMtIIa EMSA probe was made by annealing the complementary oligonucleotides 5'-GTG AAC TGA CCG CCC GCG GCC CGT GTG CAG A-3' and 5'-TCA CTC TGC ACA CGG GCC GCG GGC GGT CAG T-3'. All other double-stranded competitor oligonucleotides are as shown. Relative competition was estimated by comparing the signal obtained from the shifted complex in the presence of a varying concentrations of competitor DNA (see Figs 2 and 3 legends for concentrations).

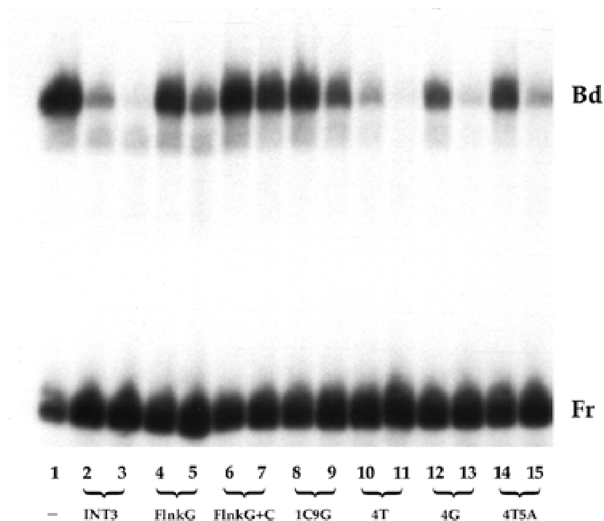


Figure 2. Representative EMSA competition analysis. Probe corresponds to the hMTIIa AP-2 binding site. Fr, free probe; Bd, bound probe. All lanes contain AP-2α ΔN165 protein derived from translation in rabbit reticulocyte lysate. The oligonucleotide competitors, listed in Table 4, are indicated below. INT3 is the wild-type competitor. The – represents the absence of specific competitor in lane 1, otherwise the even-numbered lanes contain 6.4 ng of competitor and the odd-numbered lanes contain 32 ng of competitor.

RESULTS

The PCR-based strategy utilized to identify novel AP-2 DNA binding sequences employed an affinity column containing bacterially expressed ΔN165 AP-2α protein (Fig. 1). The ΔN165 protein corresponds to the DNA binding and dimerization domains which are located in the C-terminal half of the AP-2 molecule (25). Prior to the application of randomized DNA sequences, we confirmed that this specific AP-2α affinity resin was capable of binding oligonucleotides containing the standard hMTIIa AP-2 site (data not shown). Next, double-stranded oligonucleotides with a randomized N16 core were applied to the affinity matrix in a buffer containing 100 mM KCl (see Materials and Methods). The choice of an N16 core was based on previous competition studies which indicated that this length would enable us to search for sequence preferences both within and flanking the AP-2 binding site (data not shown). Bound oligonucleotides were released from the ΔN165 column by eluting in a buffer containing a higher concentration of KCl. Three rounds of DNA application, elution and subsequent PCR were performed resulting in an ~10^6-fold enrichment (Table 1). EMSA performed using the pools of oligonucleotides derived at each round of the procedure confirmed that there was a strong selection at each round for effective AP-2 binding sequences (not shown). Oligonucleotides derived from the final round of the procedure were subcloned prior to subsequent sequence analysis.

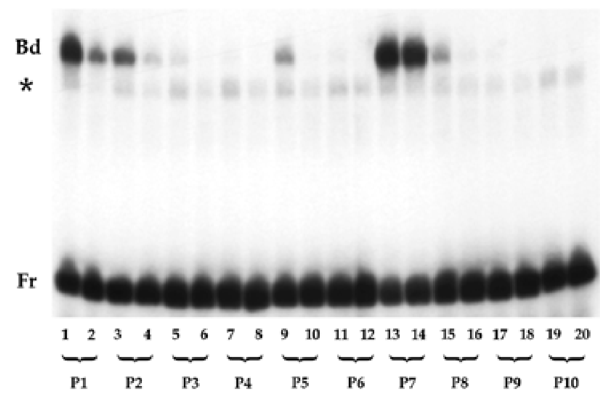
Analysis of the AP-2 binding sequences

A total of 70 independent SELEX clones were isolated and sequenced (Table 2). Strikingly, analysis of these sequences demonstrated that, although each clone was unique, certain

Table 2. AP-2α binding sites identified by Selex analysis

Clone Name	Sequence	Designation
	123 456 789	
9R	*GACC* **GCC CCA GGC** GCCTCTT*CTCG*	N3
A9	GACC **GCC TGA GGC** TCATTCCCTCG	N3
B4	GACC **GCC TAA GGC** GTAACACCTCG	N3
C3	GACC **GCC TTA GGC** ATCCTTCCTCG	N3
C4	GACC **GCC CGA GGC** GCTCTGGCTCG	N3
C11	GACC **GCC CGA GGC** TGTTCTGCTCG	N3
D1	GACC **GCC TGA GGC** GGTGATTCTCG	N3
D2	GACC **GCC TTA GGC** TATCAGGCTCG	N3
D4	GACC **GCC CTA GGC** GTTTTCCCTCG	N3
E11	GACC **GCC GCG GGC** GGTATACCTCG	N3
F7	GACC **GCC GGA GGC** TATTTTGCTCG	N3
B5	GACCC **GCC TCG GGC** TATCTGCTCG	N3
D11	GACCA **GCC TCA GGC** GACATTCTCG	N3
E9	GACCC **GCC TGC GGC** GTTTTCCTCG	N3
F10	GACCT **GCC TCA GGC** TATTCCCTCG	N3
D6	GACCTA **GCC TCA GGC** GATACCTCG	N3
A12	GACCTC **GCC TGA GGC** GTTACCTCG	N3
E7	GACCAT **GCC TTA GGC** TATTACTCG	N3
7R	GACCACC **GCC TCC GGC** GATGCTCG	N3
B10	GACCACT **GCC TCA GGC** GTTTCTCG	N3
C7	GACCAGC **GCC TGA GGC** ACAACTCG	N3
D9	GACCTAT **GCC GCG GGC** TATTCTCG	N3
A2	GACCGTATA **GCC TTA GGC** ACCTCG	N3
A6	GACCCTATC **GCC TTA GGC** TCCTCG	N3
C12	GACCTAATC **GCC TGA GGC** AACTCG	N3
F8	GACCCTGTA **GCC TGA GGC** TACTCG	N3
C9	GACCGATGTC **GCC TGG GGC** ACTCG	N3
D3	GACCGGTATA **GCC TGA GGC** GCTCG	N3
D12	GACCCGGAATAC **GCC CCA GGC** TCG	N3
C2	GACCCGTTCATA **GCC TTA GGC** TCG	N3
F1	GACCGTTAAACC **GCC TTA GGC** TCG	N3
B9	GACC **GCC TAA GGG C**ACGAAACTCG	N4
F2	GACC **GCC TTA GGG C**TTTCAGCTCG	N4
11R	GACCATT **GCC TCA GGG C**ATCCTCG	N4
E6	GACCTTCC **GCC GGA GGG C**ATCTCG	N4
E5	GACCCATC **GCC CCG GGG C**ATCTCG	N4
A1	GACCTCACTAT **GCC CGC GGG C**TCG	N4
B6′	CGAGGTATAT **GCC TTC GGG C**GGTC	N4
F6′	CGAGGTAT **GCC TAA GGG C**ATGGTC	N4
B3′	CGAGGAC **GCC TGA GGG C**ACGGGTC	N4
B8′	CGAGTAA **GCC TGA GGG C**ATTGGTC	N4
E3′	CGAGAC **GCC TCG GGG C**AATTGGTC	N4
E8′	CGAGTC **GCC TGA GGG C**ACAAGGTC	N4
8R′	CGAGC **GCC TCA GGG C**ACAATGGTC	N4
B12′	CGAGG **GCC TAA GGG C**ATAAAGGTC	N4
E2′	CGAGT **GCC TTA GGG C**GAATGGGTC	N4
F5′	CGAGC **GCC TGA GGG C**ACAAGGGTC	N4
C8′	CGAG **GCC TAA GGG C**ATACAAGGTC	N4
A4′	CGA **GCC TCC GGG C**ATACCCCGGTC	N4
C1′	CGA **GCC TTA GGG C**ATCGAATGGTC	N4
F3′	CGA **GCC TGC GGG C**ATTGGCGGGTC	N4
A3'	CGAGA **GCC CGG AGG GC**ATATGGTC	N5
B1	GACCTAT **GCC CTG GGG GC**CCCTCG	N5
A8	GACCGTTATA **GCC TGA GGG GC**TCG	N5
C6	GACCAGAGAT **GCC CTG GGG GC**TCG	N5
F9	GACCACAGAT **GCC CTC GGG GC**TCG	N5
D8	GACCTTAAT **GCC CTG AGG GGC**TCG	N6
A7'	CGAGTGAAGATA **GCC TGA GGG G**TC	V
B2'	CGAGCCTTAATA **GCC TCA GGG G**TC	V
A10'	CGAGTCAAATC **GCC CCG GGG GG**TC	V
B7'	CGAGCGGCTAT **GCC CCG AGG GG**TC	V
D5'	CGAGGGAAATA **GCC TGA GGG GG**TC	V
E12'	CGAGGGTCTAT **GCC CCC AGG GG**TC	V
E1'	CGAGTATATA **GCC CCA GGG GGG**TC	V
E4'	CGAGGAGAAT **GCC CTA AGG GGG**TC	V
E10'	CGAGTAGTAT **GCC CCC AGG GGG**TC	V
F4'	CGAGAAGGAT **GCC CTG AGG GGG**TC	V
12R'	CGAGCACAT **GCC CTG AGG GGGG**TC	V
D7'	CGAGT **GCC CGG AGG GGG**TATGGTC	V
A11	GACC **GCC TCG GGG** TATATAACTCG	V

The sequences are aligned according to the position of the binding site (bold) and spacing between GCC motifs (N3–N6). The sequence of the nucleotides flanking the random 16mer core is indicated by italics in the first example. The clone name is shown at the left. V indicates the 13 variants which contain only one GCC motif. We have also separated the core sequences of all the potential binding sites into triplets so that they can be visualized in the alternative ‘all N3’ arrangement.



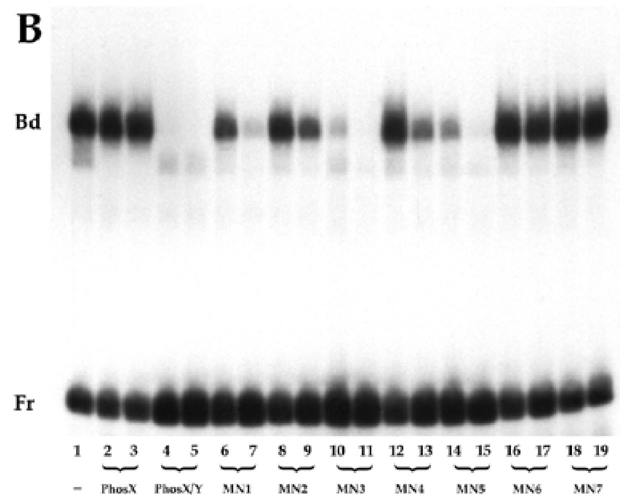


Figure 3. Missing phosphate (**A**) and missing nucleoside (**B**) analysis. In both panels, the probe corresponds to the hMTIIa AP-2 binding site. Fr, free probe; Bd, bound probe; *, non-specific complex. All lanes contain AP-2α ΔN165 protein derived from translation in rabbit reticulocyte lysate. The oligonucleotide competitors, listed in Tables 5 and 6, are indicated below. PhosX contains only one strand after self-annealing; PhosX/Y is the wild-type double-stranded competitor. The – represents the absence of specific competitor. Each pair of lanes containing a particular competitor molecule contain 7.5 ng competitor in the left lane and 37.5 ng in the right lane.

motifs appeared multiple times in the population. In particular, the most common arrangements were the 9mer sequences GCCTGAGGC and GCCTTAGGC, which occurred 11 and eight times, respectively. We also isolated multiple copies of the 10mer sequences GCCTAAGGGC (four times), GCCT-GAGGGC (four times) and GCCTTAGGGC (three times). The finding that specific motifs were present in multiple clones supports the assertion that we have selected for *bona fide* AP-2α binding sites in this analysis.

Examination of all 70 sequences revealed that at least one GCC motif was present in all instances. Indeed, 57/70 (81%) of the sequences contained two of these motifs arranged in an opposite orientation around a central region which varied from 3 to 6 nt (Table 2, N3–N6 sites). Thus, these 57 individual sequences could be classified into distinct subsets depending

Table 3. Alignment of the AP-2α binding sites

	–3	–2	–1	1	2	3	4	5	6	7	8	9	+1	+2	+3
(A) Alignment of all SELEX sequences according to orientation shown in Table 2															
A	39	20	16	0	0	0	0	5	45	9	0	0	5	28	6
G	9	10	2	70	0	0	4	24	16	61	70	39	31	16	10
C	11	22	32	0	70	70	21	22	9	0	0	31	20	16	17
T	11	18	20	0	0	0	45	19	0	0	0	0	14	10	37
	N	N	A/C/T	G	C	C	g/C/T	N	A/G/c	a/G	G	G/C	N	N	N
(B) Alignment of standard N3 sites (GCC N3 GGC) using data obtained from both orientations															
A	38	9	19	0	0	0	0	9	48	0	0	0	9	20	5
G	9	12	0	62	0	0	5	22	9	62	62	0	34	21	10
C	10	21	34	0	62	62	9	22	5	0	0	62	0	12	9
T	5	20	9	0	0	0	48	9	0	0	0	0	19	9	38
	N	N	A/C/T	G	C	C	g/C/T	N	A/G/c	G	G	C	A/G/T	N	N

upon their axis of symmetry (Tables 2 and 3). The most common arrangement, found in 31 sequences, was a pair of dyad symmetrical GCC triplets separated by three nucleotides (N3 sites). A further 20 sequences had 4 bp separating the GCC inverted repeats (N4 sites). A smaller number of sequences displayed an N5 or N6 organization (five and one clones, respectively). The remaining 13 clones contained only one GCC motif (Table 2, V sites). These latter sequences could be considered as variants of an N3 or N4 organization which contain the triplet GGG in the downstream position frequently occupied by GGC.

We also noted a second way to view the alignment of the SELEX sequences (Table 2). In this alternative configuration, all the SELEX sequences could be treated as variants of the N3 arrangement. As shown in Table 2, when we used one GCC for the initial alignment and then allowed for an intervening sequence of 3 nt (N3), we noticed a strong preference for the presence of particular triplet sequences in the adjacent six downstream positions. This was most striking for positions 7–9 of the alignment, where only the triplets GGC, GGG or AGG were observed. When this 'all N3' classification is used, 31 sequences still have the standard GCC N3 GGC configuration discussed above. An additional 30 sequences correspond to the arrangement GCC N3 GGG and nine clones have the arrangement GCC N3 AGG. Note that all nine clones containing a 3'-AGG could also be arranged in the configuration CCC N3 GGG if the binding site was shifted by 1 nt towards the 3' direction in these specific instances. This latter point again emphasizes that there are several ways to view how AP-2α might interact with a particular binding site in the SELEX population. Nevertheless, our analysis strongly supports the existence of an AP-2α consensus binding site, GCC N3 GGC, and indicates that several related sequences may also be recognized by AP-2α: GCC N4 GGC, GCC N3 GGG, GCC N3 AGG and CCC N3 GGG.

We next processed the data from the SELEX analysis to derive further information concerning the consensus AP-2α binding site. We designated the first three positions of the binding site as the trinucleotide GCC and calculated the frequency of each nucleotide in the surrounding sequence. Since many sequences contained two dyad symmetrical GCC motifs, the decision concerning which orientation to use for the alignment and subsequent data processing was somewhat arbitrary. Therefore, we analyzed the data from each orientation individually or from both orientations combined. We found that these various alignments produced the same conclusions concerning important nucleotide sequences within the AP-2α binding site (data not shown). For the sake of clarity we have only included two of these analyses in Table 3. First, in Table 3A, we have presented summary data for particular nucleotide positions based on the alignment of the 70 sequences shown in Table 2. Although this method of analysis highlights the presence of the alternative sequence arrangements GGC or GGG at positions 7–9, it also masks some of the underlying symmetry of the AP-2α binding site. Therefore, in Table 3B, we have also summarized information obtained from the 31 standard N3 sequences incorporating data derived from both strands of these symmetrical sequences (62 total sequences).

Examination of the alignments shown in Table 3 indicates that when the invariant trinucleotide GCC was employed as reference point, other nucleotide positions displayed marked sequence preferences. This observation demonstrates the quality of our alignment and strengthens the conclusion that we have isolated functional AP-2α DNA binding sites. Most notably, a guanine was always present at position 8 of the alignment, suggesting that this G:C base pair forms a critical point of interaction between AP-2α and DNA. A further interesting feature was that the position prior to the GCC was rarely occupied by a guanine. This finding indicates that the presence of a guanine nucleotide at the –1 location may be unfavorable for AP-2α binding.

There was a also a preference for particular nucleotides to be located at positions 4–6 between the inverted GCC motifs. Specifically, a thymidine was preferred in position 4 and an adenine in position 6. In fact, 19/31 sequences contain the symmetrical arrangement of these two bases in the central 3 bp of the binding site as the trinucleodides TGA/TCA or TTA/TAA. Together, the sequences GCCTGAGGC (11 incidences) and GCCTTAGGC (eight incidences) were the most commonly found in the total selection pool. Guanine or cytosine

could also be tolerated at either position 4 or 6; however, we found no incidence of adenine at position 4 (or the corresponding thymine at position 6). The data also indicate that there may be a modest preference for a G or C nucleotide in the middle position (position 5) of the binding site, since this occurs in 22/31 (71%) of the sequences. Examination of the total SELEX population shown in Tables 2 and 3 indicates that the conclusions we have derived from analysis of the standard N3 binding sites, concerning sequence preferences in positions 4 and 6, can also be applied to the remaining SELEX sequences.

Mutagenesis of the AP-2 binding site

Because we employed stringent selection criteria in the derivation of the SELEX sequences, it was possible that some weaker AP-2α DNA binding sites were discarded in earlier rounds of the procedure. Therefore, we also utilized an EMSA competition analysis both to identify further related sequence arrangements that could function as AP-2α binding sites and also to test several of the predictions of our SELEX analysis. We began with a synthetic oligonucleotide that contained a standard N3 AP-2 binding site (INT3, Table 4). In EMSA studies this INT3 sequence competed effectively against the well-characterized hMtIIa AP-2 binding site (data not shown). Next, a series of oligonucleotides based on INT3 were utilized to assess the importance of particular AP-2 binding site nucleotides. Representative results from this analysis in which we systematically substituted each nucleotide in one half of the binding site are shown in Figure 2 and the data are summarized in Table 4. We initially examined the importance of flanking sequences, in particular how a guanosine at position –1 would influence DNA binding. The presence of a guanosine at the –1 position on one side of the binding site reduced DNA binding ~10-fold (FlnkG), while the equivalent alteration at both flanking positions caused a greater than 25-fold drop in competition (FlnkG+C; Fig. 2 and Table 4). Thus, in agreement with the conclusions of the SELEX analysis, a guanosine at the –1 position is detrimental for DNA binding.

A second prediction from the SELEX study was that guanosine, and to a lesser extent cytosine, were prefered at the first position of the binding site. The EMSA experiments supported this conjecture since an oligonucleotide in which the first GCC was replaced by either ACC (1A) or TCC (1T) produced molecules which were 10- to 20-fold less effective at competition (Table 4). In contrast, alteration of this sequence to CCC (1C) reduced competition only slightly. The 1C oligonucleotide, which is an effective competitor, conforms to a consensus sequence CCC N3 GGC, in which only one GCC motif is present and the other is replaced by CCC. We next tested how the replacement of both dyad symmetrical GCC motifs by CCC would affect AP-2α binding (Fig. 2 and Table 4, 1C9G). This analysis revealed that AP-2α could bind to such a CCC N3 GGG sequence motif, albeit at a greatly reduced level of competition (~20-fold). We also performed competition studies using a second DNA oligonucleotide, SY2-35, that matched the CCC N3 GGG sequence arrangement (5'-CGA CCC CCC CGG GGG GGT CG-3'). We found that this second sequence could also compete for AP-2α binding (data not shown).

Substitutions at positions 2 or 3 of the binding site, which affected the conserved cytosines, were not well tolerated (Table 4). Only the replacement of cytosine by thymidine in the third position (GCT;3T) produced a molecule capable of significant competition. Taken together with the SELEX analysis these studies reinforce the critical nature of the GCC or CCC motif for AP-2α binding. Strong sequence preferences were also apparent in the central three nucleotides of the N3 motif. Most importantly, the presence of an adenosine at position 4 produced a very weak competitor (4A) and this sequence organization was also absent from the SELEX population. Finally, we noted that the presence of an A:T base pair at the center of N3 symmetry also produced an ~2- to 3-fold weaker competitor oligonucleotide (5A and 4T5A). This observation is consistent with the slight preference that we observed for a G:C nucleotide at the central position within the SELEX pool.

Table 4. Summary of N3 EMSA competition data

oligo	sequence	competition
	1 234 567 89	
INT3	5'- TGA CC**G CCC GGG GC**A CGT GT -3'	++++
FlnkG	5'- TGA CGG CCC GGG GCA CGT GT -3'	++
FlnkG+C	5'- TGA CGG CCC GGG GCC CGT GT -3'	-
1C	5'- TGA CCC CCC GGG GCA CGT GT -3'	+++
1C9G	5'- TGA CCC CCC GGG GGA CGT GT -3'	+
1A	5'- TGA CCA CCC GGG GCA CGT GT -3'	++
1T	5'- TGA CCT CCC GGG GCA CGT GT -3'	+
2A	5'- TGA CCG ACC GGG GCA CGT GT -3'	-
2T	5'- TGA CCG TCC GGG GCA CGT GT -3'	-
2G	5'- TGA CCG GCC GGG GCA CGT GT -3'	-
3A	5'- TGA CCG CAC GGG GCA CGT GT -3'	-
3T	5'- TGA CCG CTC GGG GCA CGT GT -3'	++
3G	5'- TGA CCG CGC GGG GCA CGT GT -3'	-
4A	5'- TGA CCG CCA GGG GCA CGT GT -3'	++
4T	5'- TGA CCG CCT GGG GCA CGT GT -3'	++++
4G	5'- TGA CCG CCG GGG GCA CGT GT -3'	+++
5A	5'- TGA CCG CCC AGG GCA CGT GT -3'	+++
4T5A	5'- TGA CCG CCT AGG GCA CGT GT -3'	+++

The INT3 sequence corresponds to a wild-type N3 consensus AP-2α site, highlighted in bold. In the other competitors, the nucleotide alterations are underlined. The ability to compete compared to INT3 is indicated on the right and ranges from equivalent (++++) to none (–) detected in the EMSA competition experiments.

Missing phosphate contact analysis

The SELEX analysis suggested that several related configurations could act as an AP-2α binding site and revealed potentially important contacts between the protein and specific

nucleotides. We next determined how other components of the binding site contributed to AP-2α protein:DNA contact. One assay that can be utilized to study protein:DNA interaction, specifically the importance of particular phosphate residues, is ethylation interference. However, analysis of the AP-2α binding site by this technique is complicated by the presence of multiple guanine residues, which are also subject to ethylation (data not shown). Therefore, we employed an alternative approach which was developed from the observation that a functional catabolite activator protein DNA binding site can be produced by the interaction of two modular half-sites (31). Our 'missing phosphate' strategy relied upon making one contiguous DNA oligonucleotide which was then annealed with two oligonucleotides that together correspond to the opposite strand. This procedure results in a single-stranded break in the sugar–phosphate backbone of the DNA. Based on the SELEX studies described above, we designed the contiguous strand to have a favorable AP-2α binding motif, but with non-complementary flanking sequences and and an asymmetric core at the center of the binding site, 5'-TCG-3'. Since it is asymmetric, this single-stranded PhosX oligonucleotide did not produce an effective competitor molecule when self-annealed (Fig. 3). However, when annealed with its equivalent complementary sequence, PhosY, it generated a strong competitor (Fig. 3 and Table 5, PhosX/Y). We were then able to anneal PhosX with pairs of oligonucleotides which together corresponded to the PhosY sequence. This yielded a series of competitor molecules that lacked individual phosphates within the core binding site. In general, we observed that the presence of the single-stranded breaks did not produce a significant drop in competition (Fig. 3 and Table 5), indicating that fully functional binding sites can be generated by this triple annealing approach. However, a single-stranded break located between nt 6 and 7 of the binding site severely reduced the ability of the oligonucleotide to compete (P7; Table 5). In addition, single-stranded breaks positioned on either side of nucleotide 1 were also detrimental, reducing competition by 5- to 10-fold (P1 and P2; Table 5).

The same series of oligonucleotides used to produce a nick corresponding to a single phosphate can also be used to produce a defined set of molecules with a gap of one nucleoside. We used this missing nucleoside strategy to probe further aspects of AP-2α protein:DNA interaction. The removal of nucleosides at positions 1, 3 or 5 of the binding site were well tolerated, reducing competition by only 5- to 10-fold (MN1, MN3 amd MN5; Fig. 3B and Table 6). In contrast, removal of nucleosides 2, 4 or 6–9 greatly reduced or abolished binding (Table 6, Fig. 3B and data not shown). When compared with the missing phosphate data, these findings suggest that important backbone or base contacts may occur in the AP-2α binding site at position 8 and, to a lesser extent, at positions 4 and 9. With respect to positions 6 and 7, the drop in binding ability may be associated with removal of the intervening phosphate. Similarly, the reduction caused by the removal of nucleotide 1 may be accounted for by the absence of the two flanking phosphates.

Table 5. Summary of N3 missing phosphate competition data

Oligo name	Core sequence	Competition
PhosX/Y	5- gCCCgAggC -3	++++
P1	5- v gCCCgAggC -3	+
P2	5- g v CCCgAggC -3	++
P3	5- gC v CCgAggC -3	+++
P4	5- gCC v CgAggC -3	++++
P5	5- gCCC v gAggC -3	++
P6	5- gCCCg v AggC -3	++++
P7	5- gCCCgA v ggC -3	-
P8	5- gCCCgAg v gC -3	++
P9	5- gCCCgAgg v C -3	++++
P10	5- gCCCgAggC v -3	++++

The central portion of the wild-type N3 sequence, PhosX/Y, is shown at the top. Chevrons indicate the positions of the single-strand breaks in other molecules. The unbroken bottom strand is not shown. Relative competition is shown on the right in comparison to PhosX/Y.

DISCUSSION

We have utilized SELEX and competition analyses to characterize the nature of the AP-2α DNA binding site. The data reveal that the organization and sequence requirements for AP-2α DNA binding are more complicated than previously described (25). One conclusion that can be drawn from our studies is that the binding site is larger than was apparent from the analysis of the SV40 and hMtIIa AP-2 binding sites alone. It is now apparent that sequence preferences for AP-2α binding span a region of 11–12 nt. The 'typical' binding site contains dyad symmetrical XGCC motifs separated by 3 or 4 nt (where X represents A, C or T). The reduction in relative binding affinity observed when a G occurs in the –1 position flanking the GCC motif was unexpected, since this arrangement occurs naturally in several instances, including hMtIIa (Table 7).

Based on the SELEX studies we postulate that there is an inherent flexibility in the AP-2α protein:DNA complex. Almost half of all the SELEX sequences (31/70) fit into a dyad symmetrical N3 format represented by the 'standard N3' site, GCC N3 GGC. Twenty of the remaining sequences fit into an alternative arrangement that can be regarded in one of two ways, either as GCC N4 GGC or GCC N3 GGG. We postulate

Table 6. Summary of N3 missing nucleoside competition data

Oligo name	Core sequence	Competition
PhosX/Y	5- AA gCC CgA ggC gA -3	++++
MN1	5- AA XCC CgA ggC gA -3	++
MN2	5- AA gXC CgA ggC gA -3	+
MN3	5- AA gCX CgA ggC gA -3	+++
MN4	5- AA gCC XgA ggC gA -3	+
MN5	5- AA gCC CXA ggC gA -3	+++
MN6	5- AA gCC CgX ggC gA -3	-
MN7	5- AA gCC CgA XgC gA -3	-
MN8	5- AA gCC CgA gXC gA -3	-
MN9	5- AA gCC CgA ggX gA -3	+

The central portion of the wild-type N3 sequence, PhosX/Y, is shown at the top. Chevrons indicate the positions of the single-strand breaks equivalent to a single nucleoside, represented by an X in other molecules. The unbroken bottom strand is not shown. Relative competition is shown on the right in comparison to PhosX/Y.

Table 7. Alignment of several naturally occurring AP-2 binding sites

N3 sites	
hMtIIa	5′- CGG **GCC** GCG **GGC** GGT -3′
E-cadherin EIII	5′- GGC **GCC** CGA **GGC** GAG -3′
MMTV LTR	5′- TTT **GCC** TGG **GGC** TAT -3′
R-FABP	5′- GTT **GCC** GTG **GGC** GGC -3′
mouse H-2K^b	5′- GAA **GCC** CAG **GGC** TGG -3′
IGFBP-5	5′- GCA **GCC** AGG **GGC** CGG -3′
c-erbB-2	5′- ACG **GCT** GCA **GGC** AAC -3′
SV40	5′- AAG **TCC** CCA **GGC** TCC -3′
N4 sites	
K14	5′- GTA **GCC** TGC **AGG C**CC -3′
ENK	5′- ATC **GCC** GGC **GGG C**TG -3′
c-myc	5′- AGG **GCC** GGT **GGG C**GG -3′
Variant sites	
E-cadherin EII	5′- GT**G CCC** TGA **GGG GGG** -3'
ER	5′- CT**G CCC** TGC **GGG G**AC -3'
K3	5′- G**CC CCC** TGC **AGG G**TA -3'
XK81A1	5′- GTA **GCC** TCA **GGG** TGT -3'

The binding sites are derived from: hMtIIa, the human metallothionein IIa promoter, nt –168 to –182 (20,42); mouse E-cadherin enhancer EII and III elements, nt 520–534 and 562–576, respectively (19); MMTV LTR, the mouse mammary tumor virus long terminal repeat, nt –1024 to –1010 (43); R-FABP, chick retinal fatty acid-binding protein promoter, nt –78 to –64 (44); mouse H-2k^b promoter, nt –183 to –169 (45); IGFBP-5, human insulin-like growth factor-binding protein-5 promoter, nt –49 to –35 (46); human c-erbB-2 promoter, nt –223 to –209 (18); SV40, nt 244–229 (20,42); K14, human keratin K14 promoter, nt –234 to –219 (47); ENK, human proenkephalin promoter, nt –67 to –81 (48); human c-myc, nt 620–606 (42); ER, human estrogen receptor promoter, nt 186–200 (49); K3, rabbit keratin K3 promoter, nt –201 to –187 (50); XK81A1, *Xenopus* keratin XK81A1, nt –150 to –164 (51).

that these latter 20 sequences are fundamentally different from the standard N3 site due to specific sequence preferences at position 10 of the binding site in both instances. None of the 31 standard N3 sites has a cytosine at position 10 following the GGC triplet. However, if GGG occurs at positions 7–9 then the majority of the sequences (20/30) have a cytosine in position 10. This observation implies that the AP-2α protein may be able to adjust to accommodate both of these alternative configurations. The flexibility required to contact dyad symmetrical GCC triplets separated by 3 or 4 nt could be provided either by the dimerization domain of AP-2α protein or by distortion of the DNA binding site. This flexibility might also extend to N5 and N6 sites. However, given the limited number of these N5 and N6 sequences obtained we postulate that these sites may represent fortuitous sequence variants of the N3/N4 configurations. Further studies will be required to determine if AP-2α can induce DNA conformational changes and/or whether the AP-2 dimer has an inherent flexibility which may enable it to interact with either an N3 or N4 site.

A third important conclusion of our analysis is that there are particular sequence preferences at positions 4–6 of the standard N3 binding site. Specifically, there is a significant reduction in the relative affinity for AP-2α when the binding site contains an adenosine in position 4 or a thymidine in position 6. We also noted that although the EMSA studies indicated that any nucleotide could occupy each of the middle three positions of the binding site, specific triplet combinations often occurred in this location. In particular, the dyad symmetrical arrangements TGA/TCA or TAA/TTA were present at these positions in over half of the SELEX population. Other combinations were less common, especially CAG, CTC, CAA, CAC, GTA and GAA, in which an A or T in the middle position is placed in a context other than TTA/TAA. Competition studies with sequences containing some of these latter combinations (Table 4: 5A, CAG; 4T5A, TAG) confirmed that they had a reduced AP-2α DNA binding ability. The molecular

basis for these sequence preferences will likely require a detailed structural determination of the AP-2α protein in association with its cognate DNA binding site.

We further analyzed the specificity of AP-2α protein:DNA contact using a combination of missing phosphate and missing nucleoside analyses. A modified version of the missing nucleoside approach, which relies on hydroxyl radical footprinting to generate a gap of one nucleoside, has previously been described for analysis of the λ repressor–O_R1 complex (32). Similarly, the potential influence of a particular phosphate on protein:DNA contact can be examined by an ethylation interference experiment (33). In this latter technique *N*-ethyl-*N*-nitrosourea (ENU) is used to attach an ethyl group onto the phosphate residues of the DNA backbone. The ability of this additional moiety to interfere with protein:DNA interaction is then assessed. However, ENU efficiently alkylates guanines too and can also modify the other bases at a variety of positions (34). The lack of specificity of ENU, coupled with the production of multiple bands derived from the same nucleoside following subsequent cleavage, can greatly complicate data interpretation. The methodology we have employed provides an alternative approach to these chemical modification techniques and also enables phosphate contacts to be studied by a method which complements ethylation interference.

We should note that the introduction of single-stranded breaks can alter the physical properties of a DNA molecule (35–40). In particular, a gap of 1 nt can alter the flexibility of DNA and may also disrupt B-form structure. Therefore, any reduction in protein:DNA affinity caused by a single-stranded break may indicate that important contacts have been lost or that the structure and/or flexibility of the DNA has been altered so that protein binding is compromised. Nevertheless, a comparison of the data obtained concerning the cytosines located at positions 2 and 3 of the binding site illustrates the potential utility of missing nucleoside analysis for studying protein:DNA interaction. Substitutions involving the C:G base pair at the second position of the binding site are not tolerated in competition studies. Furthermore, the cytosine itself may be important for AP-2α base contact at this location, since the actual removal of this nucleoside results in a significant drop in competition. Cytosine was also preferred in position 3. However, in this instance, the vital protein:DNA contacts are presumably occurring with the complementary base on the opposite strand because complete removal of the cytosine had little influence on AP-2α binding.

We had previously examined AP-2 protein:DNA contact using a missing contact analysis, a chemical modification technique in which individual bases are removed from the DNA without breaking the sugar–phosphate backbone (25). A comparison between these previous missing contact studies performed on the SV40 and hMtIIa AP-2α binding sites and the current missing nucleoside analysis indicates that the relative importance of similar nucleotide positions is conserved in all three instances. The similarity of the data obtained with these two different assays confirms the efficacy of our missing nucleoside approach and strengthens our conclusions concerning the critical AP-2α binding site contacts.

The missing phosphate analysis revealed further information concerning the requirements for AP-2α DNA recognition. In particular, the presence of a gap located between positions 6 and 7 of an N3 site abolished AP-2α DNA binding. The sensitivity of the AP-2α site to the presence of a single-stranded gap just prior to the downstream GGC motif could indicate that this removes a critical phosphate contact for the protein:DNA interaction. Indeed, the location of this phosphate would position it on the same side of the DNA helix as the essential residues identified by missing contact and methylation interference analyses (25). One plausible mechanism by which this phosphate plays such an important role in AP-2α recognition is by 'buttressing' otherwise weak interactions between the protein sidechains and a specific base, as occurs in the *trp* repressor complex (41). An alternative possibility is that the AP-2α protein requires an alteration in DNA conformation that is prevented by the presence of a single-stranded break just prior to the GGC motif. One argument for an influence of AP-2 on DNA conformational changes, such as DNA bending or twisting, is our finding that the AP-2α molecule may be able to recognize the dyad symmetrical GCC motifs in both an N3 and N4 context.

Recently, an independent AP-2α binding site selection study proposed that the sequence 5'-TAG AAA GNY CYN G-3' was an AP-2α consensus binding site (27). Our SELEX and competition studies do not reveal any evidence of this potential consensus sequence. We believe that the discrepancy between the two studies results primarily from the omission of non-specific competitor DNA from their binding site selection, but also from the significant difference in sample size between the two studies (four clones in the previous study as opposed to the 70 analyzed above). The reliability of our new AP-2α consensus sequence is also supported by the finding that several of our SELEX sequences match known AP-2 binding sites (Table 7). Indeed, the combination of SELEX and competition studies we have employed shows that the ability of any particular sequence to bind AP-2α will result from several considerations, including the GCC motifs, the number and sequence of the intervening nucleotides and the flanking sequence. In this regard, our SELEX pool did not contain several variants of the AP-2 binding site which have been previously characterized. For example, we did not recover molecules in which the GCC motif was represented by GCT or TCC, as occurs in the c-erbB-2 and SV40 AP-2 sites, respectively, nor sites in which an A was present in position 4, as occurs in the IGFBP-5 promoter (Table 7). Competition studies indicate that these are weaker AP-2 binding sites which therefore may be unable to withstand the high ionic strengths employed in our SELEX protocol. Nevertheless, it is clear that such weaker AP-2 binding sites, which deviate in one or more positions from the consensus sequence, can act as critical *cis*-regulatory elements. In conclusion, our current analysis provides criteria for predicting the ability of a particular sequence to act as an effective AP-2α binding site and will assist in the identification of target genes for this critical transcription factor. It should also be noted that all known AP-2 family members are highly related in their DNA binding and dimerization domains (1,2,4). Therefore, it is likely that the information obtained from this study will be applicable to the related AP-2β and AP-2γ proteins.

ACKNOWLEDGEMENTS

We thank Mark Biggin for critical reading of the manuscript, William Russ, Paul Sigler, Thomas A. Steitz and members of the Williams laboratory for advice and assistance, the Keck

Facility for oligonucleotide synthesis and DNA sequence analysis, and Lisa Chan, Mitch Zollinger and Jose Tormo for help with protein synthesis. This research was supported by National Institutes of Health grants CA77833 and DE12728 (T.W.). J.A.I. was supported by a NIH post-doctoral fellowship GM17645 and by NIH grant GM22778 (T.A.S.).

REFERENCES

1. Bosher,J.M., Totty,N.F., Hsuan,J.J., Williams,T. and Hurst,H.C. (1996) *Oncogene*, **13**, 1701–1707.
2. Chazaud,C., Oulad-Abdelghani,M., Bouillet,P., Décimo,D., Chambon,P. and Dollé,P. (1996) *Mech. Dev.*, **54**, 83–94.
3. Mitchell,P.J., Timmons,P.M., Hebert,J.M., Rigby,P.W.J. and Tjian,R. (1991) *Genes Dev.*, **5**, 105–119.
4. Moser,M., Imhof,A., Pscherer,A., Bauer,R., Amselgruber,W., Sinowatz,F., Hofstädter,F., Schüle,R. and Buettner,R. (1995) *Development*, **121**, 2779–2788.
5. Moser,M., Pscherer,A., Roth,C., Becker,J., Mucher,G., Zerres,K., Dixkens,C., Weis,J., Guay-Woodford,L., Buettner,R. and Fassler,R. (1997) *Genes Dev.*, **11**, 1938–1948.
6. Nottoli,T., Hagopian-Donaldson,S., Zhang,J., Perkins,A. and Williams,T. (1998) *Proc. Natl Acad. Sci. USA*, **95**, 13714–13719.
7. Schorle,H., Meier,P., Buchert,M., Jaenisch,R. and Mitchell,P.J. (1996) *Nature*, **381**, 235–238.
8. West-Mays,J.A., Zhang,J., Nottoli,T., Hagopian-Donaldson,S., Libby,D., Strissel,K.J. and Williams,T. (1999) *Dev. Biol.*, **206**, 46–62.
9. Zhang,J., Hagopian-Donaldson,S., Serbedzija,G., Elsemore,J., Plehn-Dujowich,D., McMahon,A.P., Flavell,R.A. and Williams,T. (1996) *Nature*, **381**, 238–241.
10. Means,A.L. and Gudas,L.J. (1995) *Annu. Rev. Biochem.*, **64**, 201–233.
11. Lüscher,B., Mitchell,P.J., Williams,T. and Tjian,R. (1989) *Genes Dev.*, **3**, 1507–1517.
12. Oulad-Abdelghani,M., Bouillet,P., Chazaud,C., Dollé,P. and Chambon,P. (1996) *Exp. Cell Res.*, **225**, 338–347.
13. Shen,H., Wilkie,T., Ashique,A.M., Narvey,M., Zerucha,T., Savino,E., Williams,T. and Richman,J.M. (1997) *Dev. Biol.*, **188**, 248–266.
14. Williams,T., Admon,A., Lüscher,B. and Tjian,R. (1988) *Genes Dev.*, **2**, 1557–1569.
15. Zeng,Y.-X., Somasundaram,K. and El-Deiry,W.F. (1997) *Nature Genet.*, **15**, 78–82.
16. Jean,D., Gershenwald,J.E., Huang,S., Luca,M., Hudson,M.J., Tainsky,M.A. and Bar-Eli,M. (1998) *J. Biol. Chem.*, **273**, 16501–16508.
17. Turner,B.C., Zhang,J., Gumbs,A.A., Maher,M.G., Kaplan,L., Carter,D., Glazer,P.M., Hurst,H.C., Haffty,B.G. and Williams,T. (1998) *Cancer Res.*, **58**, 5466–5472.
18. Bosher,J.M., Williams,T. and Hurst,H.C. (1995) *Proc. Natl Acad. Sci. USA*, **92**, 744–747.
19. Hennig,G., Lowrick,O., Birchmeier,W. and Behrens,J. (1996) *J. Biol. Chem.*, **271**, 595–602.
20. Mitchell,P.J., Wang,C. and Tjian,R. (1987) *Cell*, **50**, 847–861.
21. Fini,M.E., Bartlett,J.D., Matsubara,M., Rinehart,W.B., Mody,M.K., Girard,M.T. and Rainville,M. (1994) *J. Biol. Chem.*, **269**, 28620–28628.
22. Somasundaram,K., Jayaraman,G., Williams,T., Moran,E., Frisch,S. and Thimmapaya,B. (1996) *Proc. Natl Acad. Sci. USA*, **93**, 3088–3093.
23. Batsché,E., Muchardt,C., Behrens,J., Hurst,H.C. and Crémisi,C. (1998) *Mol. Cell. Biol.*, **18**, 3647–3658.
24. Gaubatz,S., Imhof,A., Dosch,R., Werner,O., Mitchell,P., Buettner,R. and Eilers,M. (1995) *EMBO J.*, **14**, 1508–1519.
25. Williams,T. and Tjian,R. (1991) *Genes Dev.*, **5**, 670–682.
26. Williams,T. and Tjian,R. (1991) *Science*, **251**, 1067–1071.
27. Gee,M.S., Sarkisian,C.J. and El-Deiry,W.S. (1998) *Biochem. Biophys. Res. Commun.*, **243**, 307–316.
28. Rosenberg,A.H., Lade,B.N., Chui,D., Lin,S.-W., Dunn,J.J. and Studier,F.W. (1987) *Gene*, **56**, 125–135.
29. Ekker,S.C., Young,K.E., von Kessler,D.P. and Beachey,P.A. (1991) *EMBO J.*, **10**, 1179–1186.
30. Creaser,P., D'Argenio,D.A. and Williams,T. (1996) *Nucleic Acids Res.*, **24**, 2597–2605.
31. Schultz,S.C., Shields,G.C. and Steitz,T.A. (1990) *J. Mol. Biol.*, **213**, 159–166.
32. Dixon,W.J., Hayes,J.J., Levin,J.R., Weidner,M.F., Dombroski,B.A. and Tullius,T.D. (1991) *Methods Enzymol.*, **208**, 380–413.
33. Wissmann,A. and Hillen,W. (1991) *Methods Enzymol.*, **208**, 365–379.
34. Sun,L. and Singer,B. (1975) *Biochemistry*, **14**, 1795–1802.
35. Koudelka,G.B., Harbury,P., Harrison,S.C. and Ptashne,M. (1988) *Proc. Natl Acad. Sci. USA*, **85**, 4633–4637.
36. Le Cam,E., Fack,F., Ménissier-de Murcia,J., Cognet,J.A.H., Barbin,A., Sarantoglou,V., Révet,B., Delain,E. and de Murcia,G. (1994) *J. Mol. Biol.*, **235**, 1062–1071.
37. Aymami,J., Coll,M., van der Marel,G.A., van Boom,J.H., Wang,A.H.-J. and Rich,A. (1900) *Proc. Natl Acad. Sci. USA*, **87**, 2526–2530.
38. Giraud-Panis,M.-J., Toulmé,F., Maurizot,J.C. and Culard,F. (1992) *J. Biomol. Struct. Dyn.*, **10**, 295–309.
39. Mills,J.B., Cooper,J.P. and Hagerman,P.J. (1994) *Biochemistry*, **33**, 1797–1803.
40. Roll,C., Ketterlé,C., Faibis,V., Fazakerley,G.V. and Boulard,Y. (1998) *Biochemistry*, **37**, 4059–4070.
41. Sigler,P.B. (1992) In McKnight,S.L. and Yamamoto,K.R. (eds), *Transcriptional Regulation*. Cold Spring Harbor Laboratory Press, Plainview, NY, pp. 475–499.
42. Imagawa,M., Chiu,R. and Karin,M. (1987) *Cell*, **51**, 251–260.
43. Mellentin-Michelotti,J., John,S., Pennie,W.D., Williams,T. and Hager,G.L. (1994) *J. Biol. Chem.*, **269**, 31983–31990.
44. Bisgrove,D.A., Monckton,E.A. and Godbout,R. (1997) *Mol. Cell. Biol.*, **17**, 5935–5945.
45. Israël,A., Le Bail,O., Hatat,D., Piette,P., Kieran,M., Logeat,F., Wallach,D., Fellous,M. and Kourilsky,P. (1989) *EMBO J.*, **8**, 3793–3800.
46. Duan,C. and Clemmons,D.R. (1995) *J. Biol. Chem.*, **270**, 24844–24851.
47. Leask,A., Byrne,C. and Fuchs,E. (1991) *Proc. Natl Acad. Sci. USA*, **88**, 7948–7952.
48. Hyman,S.E., Comb,M., Pearlberg,J. and Goodman,H.M. (1989) *Mol. Cell. Biol.*, **9**, 321–324.
49. McPherson,L.A., Baichwal,V.R. and Weigel,R.J. (1997) *Proc. Natl Acad. Sci. USA*, **94**, 4342–4347.
50. Chen,T.-T., Wu,R.-L., Castro-Munozledo,F. and Sun,T.-T. (1997) *Mol. Cell. Biol.*, **17**, 3056–3064.
51. Snape,A.M., Winning,R.S. and Sargent,T.D. (1991) *Development*, **113**, 283–293.