Nucleic Acids Research, Vol. 20, No. 12 3043–3049

Extension of the DNA binding consensus of the chicken c-Myb and v-Myb proteins

Kathleen Weston
Institute of Cancer Research, Chester Beatty Laboratories, 237 Fulham Road, London SW3 6JB, UK

Received March 17, 1992; Revised and Accepted May 26, 1992

ABSTRACT

The chicken c-*myb* gene and the v-*myb* oncogene transduced by avian myeloblastosis virus (AMV) encode DNA binding transcription activators. The DNA binding domain of AMV v-Myb displays a number of amino acid changes relative to c-Myb; v-Myb proteins in which one or more of three crucial residues in the DNA binding domain are mutated to resemble the c-Myb sequence display altered transformation phenotypes. In order to establish whether the spectrum of DNA binding sites which AMV v-Myb can recognise is different from that seen by chicken c-Myb, a site selection protocol was used to derive consensus binding sequences for three variant Myb proteins made *in vitro*, and also using nuclear extract from the v-*myb* transformed cell line BM2. The results show that the original consensus binding site defined for v-Myb, YAACKG, can be extended to YAACKGHH, and that this new consensus holds for both v-Myb and chicken c-Myb.

INTRODUCTION

The v-*myb* oncogene is the transforming gene of two separate avian retroviruses, avian myeloblastosis virus (AMV) and erythroblastosis virus 26 (E26); in the latter, it is fused to a second oncogene, v-*ets*. Like its cellular progenitor c-*myb*, v-*myb* encodes a DNA-binding transcription activator (1). The c-Myb DNA binding domain is at the amino terminus of the protein, and is contained within a tripartite 52 amino acid repeat distinguished by the occurrence of tryptophan residues spaced either 18 or 19 residues apart (2, 3). Computer modelling of the domain suggests that it has some similarity to the helix turn helix structure proposed for the homeobox DNA binding domain (4). Both AMV and E26 v-Myb are amino-terminally truncated relative to c-Myb, and have lost almost all of the first repeat unit, which has been demonstrated to be unnecessary for specific DNA binding (5, 6). Whilst the DNA binding domain of E26 v-Myb is identical to c-Myb, that of AMV v-Myb contains 4 point mutations. The differentiation state of v-*myb* transformed cells appears to be tightly linked to three of these point mutations. *In vitro*, v-*myb* will only transform cells of the myelomonocytic lineage. When the c-*myb* DNA binding domain is used, transformed cells resemble myeloblasts. When the AMV v-*myb* sequence is used, cells have a monoblast phenotype. Individual mutation of the three residues causes phenotypes intermediate between the two extremes (7). One explanation for this observation might be that one or more of the three residues is in a critical position in the DNA binding domain involved in site-specific recognition of DNA; mutation might subtly alter the binding site specificity of the AMV v-Myb protein and hence allow regulation of a different subset of genes, leading to differences in cell morphology after transformation. In order to test this model, a site selection protocol was used to generate binding site consensus sequences for chicken c-Myb, AMV v-Myb and a truncated mutant of v-Myb. The results show that the binding site consensus for all three proteins is slightly larger than that originally defined for v-Myb, but remains essentially the same in all cases.

MATERIALS AND METHODS

Plasmids

pT7βV: The entire AMV v-*myb* coding sequence was removed from plasmid vmybgagSP65 (1) by BamHI digestion and gel purification, and recloned into the BamHI site of a pUC12 derivative whose EcoRI and SalI sites had been destroyed by digestion, end repair and reclosure. This plasmid, pvgag, was digested with EcoRI and BamHI, and the 800bp 3′ v-*myb* sequence was gel purified. A 360bp fragment carrying the 5′ end of v-*myb* was generated by EcoRI digestion followed by partial NcoI digestion. The 5′ and 3′ ends of v-*myb* were ligated into pT7βSal (8) cut with NcoI and BamHI, to give pT7βV. pT7βV therefore contains the v-*myb* coding region linked to the human β globin initiating methionine and Kozak sequence.

pT7βXT: The v-*myb* coding region was subjected to site directed mutagenesis to delete nucleotides 830–880 (numbering as in 9), which encode amino acids 215–232 in the v-*myb* activator domain. Mutated v-*myb* was cloned into the pUC12 derivative described above to give pvgagSD11. pvgagSD11 was cut with SalI, end-repaired, and ligated with an XbaI termination linker CTAGTCTAGACTAG to give pvXT. pvXT was digested as for pvgag and cloned into pT7βSal cut with NcoI and BamHI.

pT7βC: A fragment carrying the 5′ end of the chicken c-*myb* coding region (nucleotides 30–687 encoding amino acids 2–186) was removed from pneoAMV5′SAcmyb (1) by Polymerase Chain Reaction (PCR) using primers CM2 (TCCGGCGCATG-GTGGAATTC) and CM4 (GGATCCTGCGCACGGAGACC-CCGGCA). The fragment was digested with AviII and EcoRI,

and cloned into pT7βSal first cut with NcoI and end-repaired, and then cut with EcoRI. The resulting plasmid, pT7βC5′, was then digested with EcoRI and BamHI, and ligated with a 1520bp fragment carrying the 3′ end of the c-*myb* coding region cut from pneoAMV5′SAcmyb with EcoRI and BamHI.

Proteins

In vitro transcription and translation. Plasmids pT7βXT, pT7βV and pT7βC were linearised with BamHI, and transcribed *in vitro* by standard methods (10). Capped cRNA was translated as previously described (11). The integrity of the resulting XTmyb, Vmyb and Cmyb proteins was checked by Western blotting using monoclonal antibody myb2.74 (12), or by labelling a small portion of the translation reaction using ^{35}S-methionine, followed by SDS-PAGE. All proteins were assayed for functional DNA binding activity in gel mobility shift assays using oligonucleotide 9M1/9M2 (see below) as probe.

BM2 nuclear extracts. Nuclear extract from the transformed chicken myeloid line BM2 was made essentially as described in (13), except that buffer A contained 0.23M sucrose and 0.1% NP40, buffer C contained 0.32M NaCl, and extracts were flash frozen in buffer C without dialysis. The protein concentration of the extract was 6.5mg/ml.

Oligonucleotides and probes

Site selection oligonucleotide. Oligonucleotide R76 (CAGGTCAGATCAGCGGATCCTGTCG(N)$_{26}$GAGGCGAATTCAGTGCATGTGCAGC) was double stranded using the complementary primer F (GCTGCACATGCACTGAATTCGCCTC) as previously described (14). Primers F and R (CAGGTCAGATCAGCGGATCCTGTCG) were used for PCR amplification during site selection.

Gel mobility shift probes. 9M1: CTAGGACATTATAACGGTTTTTTAGT; 9M2: CTAGACTAAAAAACCGTTATAATGTC; 9MU1: CTAGGACATTATCACGGTTTTTTAGT; 9MU2: CTAGACTAAAAAACCGTGATAATGTC; 1K: CTAGGACATTACAAC; 2K: CTAGACTAAAAAACAGTTGTAATG; 3K: CTAGACTACCACCCCGTTATAATG; 4K: CTAGACTAAAACCCCGTTATAATG; 5K: CTAGACTAAAAACCCGTTATAATG; 6K: CTAGACTAAAACACCGTTATAATG; 7K: CTAGACTAAAAATCCGTTATAATG; 8K: CTAGACTAAATGACCGTTATAATG; 9K: CTAGGACATTATAAC. Oligonucleotides were annealed in 50mM NaCl and end-labelled under standard conditions using AMV reverse transcriptase, 40μCi α-^{32}P dCTP, and 0.2mM dATP, dGTP and dTTP. Annealings: 9M1/9M2, 9MU1/9MU2, 1K/2K, 9K with 3K, 4K, 5K, 6K, 7K, 8K, and 9M2.

Site selection

Site selection was performed as described in (14), except that immune precipitation was for 1 hour rather than overnight, with 3μl of monoclonal antibody myb 2.74 (12) prebound to 30μl of Protein A Sepharose. Myb2.74 recognises an epitope in the region of Myb immediately after the amino-terminal DNA binding domain. 2μl of programmed reticulocyte lysate or BM2 nuclear extract was used per round of site selection. Selected oligonucleotides were cloned as described (14), by performing a gel mobility shift assay in the presence of 1μl myb 2.74 antibody, excising the band corresponding to the supershifted DNA:protein:antibody complex and amplifying eluted DNA by PCR. Following digestion with BamHI and EcoRI, fragments were cloned into M13mp18 or pUC18, and sequenced by standard methods using T7 DNA Polymerase (Pharmacia). During sequence analysis, a central random sequence of less than 26 bp was occasionally found; these were only scored if sequenced on both strands to eliminate the possibility of compression artefacts.

Gel mobility shift assays

Binding reactions (100mM KCl) contained 12μl buffer E (14), 2μg poly(dIdC).(dIdC), 2μg calf thymus DNA, 2μl programmed reticulocyte lysate or 4μl BM2 nuclear extract, and 0.25μg probe in a 20μl final volume. For supershifts, 1μl of myb 2.74 antibody was also added. Reactions were incubated for 30 minutes on ice, then loaded onto a 4% 30:1 crosslinked polyacrylamide gel containing 0.5×TBE and 4% glycerol. Gels were run in the cold for 3 hrs at 15V/cm, dried onto 3MM paper and autoradiographed. Several myβ-specific complexes were always observed using the V and C myb lysates, as both contained a number of prematurely terminated myb species. Extra bands evident in BM2 gel shifts were presumably due either to some degradation of the endogenous v-Myb protein, or to the presence in the extract of other proteins capable of binding the 9M1 site. As only the three slowest-migrating complexes were recognised by Myb2.74, a combination of these possibilities seems likely.

RESULTS

Myb proteins used for site selection

Four sources of myb protein were used in the binding site selection assay. XT, V and C myb proteins, diagrammed in Figure 1a, were made by translation in rabbit reticulocyte lysate. The sequences of the V and C myb proteins correspond exactly to the AMV v-Myb oncoprotein and the chicken c-Myb proto-oncoprotein respectively. XT myb is a C-terminally-truncated derivative of AMV v-Myb, which lacks the myb activator domain, and additionally binds DNA with greater affinity than either v- or c-Myb (15). As an alternative source of AMV v-Myb, nuclear extract was prepared from the AMV-transformed chicken myeloid line BM2; this line expresses large quantities of v-Myb, but no detectable c-Myb (16). Three different versions of AMV v-Myb were used for site selection in order to determine whether the source or binding affinity of the test protein had any effect on binding site recognition.

Before use for site selection, all four protein sources were tested in gel mobility assays using as probe an oligonucleotide, 9M1, which contained the specific Myb binding site defined as Site A in the chicken *mim*-1 gene (17). A typical assay is shown in Figure 1b. Complexes which could be specifically competed with an excess of cold 9M1 (lanes 2, 6, 10 and 14), but not with mutant oligonucleotide 9MU1 (lanes 3, 7, 11, and 15), and which were supershifted with the Myb-specific monoclonal antibody Myb2.74 (lanes 4, 8, 12 and 16), formed in all four cases.

Binding site selection

Site selections were performed using as starting material the double-stranded oligonucleotide R76, in which 26 bp of random sequence are flanked by two different 25 bp fixed sequences. After each round of binding, specific Myβ-DNA complexes were immunoprecipitated using antibody Myb2.74, and the DNA was eluted and amplified by PCR using primers complementary to the fixed ends of the oligonucleotide. Four rounds of site selection

were performed with each of the Myb protein sources. A representative gel mobility assay monitoring specific complex formation during the selection process is shown in Figure 2. A faint complex (lane 2), visible after the second round of site selection, is further amplified in rounds three and four (lanes 3 and 4). The round 4 complex can be specifically competed with excess cold 9M1 (lanes 5 and 6), and is supershifted by antibody (lane 7). Selection with each of the myb protein sources was similarly monitored, and DNA eluted from all four round 4 supershifted complexes was amplified, cloned and sequenced.

Compilation of selected sequences

The sequences of the round four oligonucleotides are shown in Figure 3. Examination of the sequences revealed that 209 of the total of 213 analysed contained either a 5/6 or perfect match to the originally defined v-Myb binding consensus, $^{T}/_{C}AAC^{G}/_{T}G$ (18). It therefore appears that this consensus forms a common binding motif for all the four Myb proteins assayed. Additionally, 17/43 XT sequences, 29/67 V sequences, 11/61 C sequences and 16/42 BM2 sequences contained two or more putative binding sites. Although the frequency of multiple sites was greater than might be expected by chance, and thus might reflect a genuine binding preference of the Myb proteins, there was no obvious pattern in the orientation or spacing of the sites.

Derivation of an extended Myb binding consensus

In order to determine whether Myb binding sites comprise more than the simple core motif, oligonucleotides containing a single Myb site were aligned as shown in Figure 3. Initial analysis concentrated on the single site oligonucleotides, as it was not possible to determine unequivocally which sites were genuine in oligonucleotides with two or three Myb motifs. However, comparison of the base scores for single sites only and for both single and multiple sites showed that base preferences were essentially the same for either; accordingly, consensus sequences were generated using all sites. For the analysis, 5/6 and 6/6 matches to the Myb core consensus which contained the sequence AAC at positions 2, 3 and 4 were scored as real sites. 6/6 matches to the Myb core consensus in which the outermost base was derived from the fixed flanking sequences were scored, but no fixed sequences were included in subsequent scoring. Table 1 details the base preferences of the four different Myb protein sources. Within the hexamer core, as might be predicted, there was very little variation between the different proteins, except at positions 1 and 5; at position 1, the T:C ratio of the XT, V and BM2 proteins was ~70%:24% whereas that of C-Myb was 51%:39%, and at position 5, the XT, V and BM2 proteins preferred a G residue (>82% of sequences), whereas the C protein selected G (48%), T (32%) or C (18%). Table 2 summarises the consensus sequences proposed for each of the four proteins, together with a general consensus for all four together. No base preferences on the 5′ side of the hexamer core were observed for any of the sequences, but a weak consensus could be defined for the 3′ 5bp. Although this was slightly different for each protein, very similar binding preferences were evident, the most marked being an aversion for G residues at positions +1, +2, +4 and +5.

Verification of an extended myb consensus sequence

A series of double stranded oligonucleotide probes (Figure 4a) was constructed to test the validity of the base preferences inferred from the site selection assays. Gel mobility shift assays were performed using XT, V and C proteins, and equal amounts of the nine probes. Results are shown in Figure 4b. All three proteins behaved virtually identically with respect to their preferences for specific probes. As shown in Figure 1b and lane 2 of Figure 4, probe 9M1 contains a very efficient binding site; this site can be destroyed (probe 9MU1, lane 3) by mutation of the AAC core of the myb hexamer to CAC (18). Interestingly, mutation of the Myb core from TAACGG to CAACTG appears

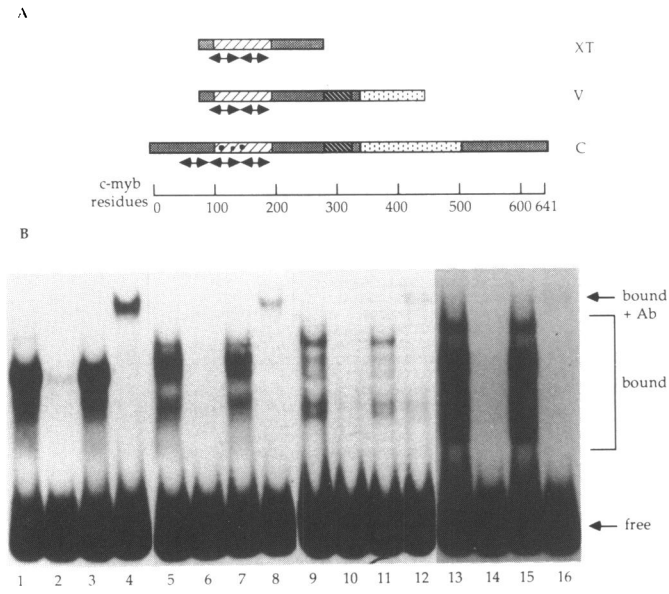

Figure 1. A. Structure of reticulocyte lysate-translated myb proteins. Arrows indicate repeat units of the DNA-binding domain. The minimal DNA binding domain (wide-striped), activator domain (dark-striped) and repressor domain (stippled) are also shown. The three dots in the DNA binding domain of c-Myb indicate amino acid changes relative to v-Myb. B. Gel mobility shift assays using oligonucleotide 9M1 as probe. Lanes 1–4: XT Myb. Lanes 5–8: v-Myb. Lanes 9–12: BM2 nuclear extract. Lanes 13–16: c-Myb. Lanes 1, 5, 9 and 13 are simple gel shifts. Lanes 2,6,10 and 14 additionally contain a 100-fold excess of cold 9M1 oligonucleotide, lanes 3,7,11 and 15 a 100-fold excess of 9MU1 (see text), and lanes 4, 8, 12 and 16 monoclonal antibody Myb2.74.

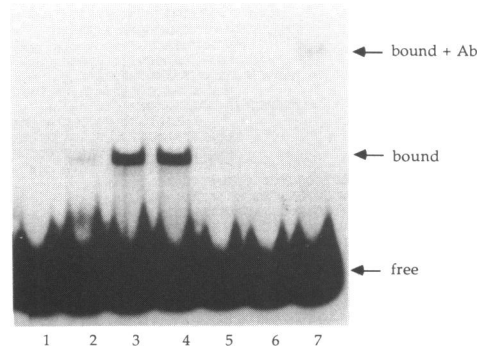

Figure 2. Gel mobility shift assay of labelled oligonucleotide pools selected using XTMyb protein. Lanes 1–4: rounds 1–4 of site selection. Lanes 5 and 6: round 4 pool with 100-fold excess of cold 9M1 oligonucleotide alone (lane 5) or together with antibody Myb2.74 (lane 6). Lane 7: round 4 pool supershifted with antibody Myb2.74.

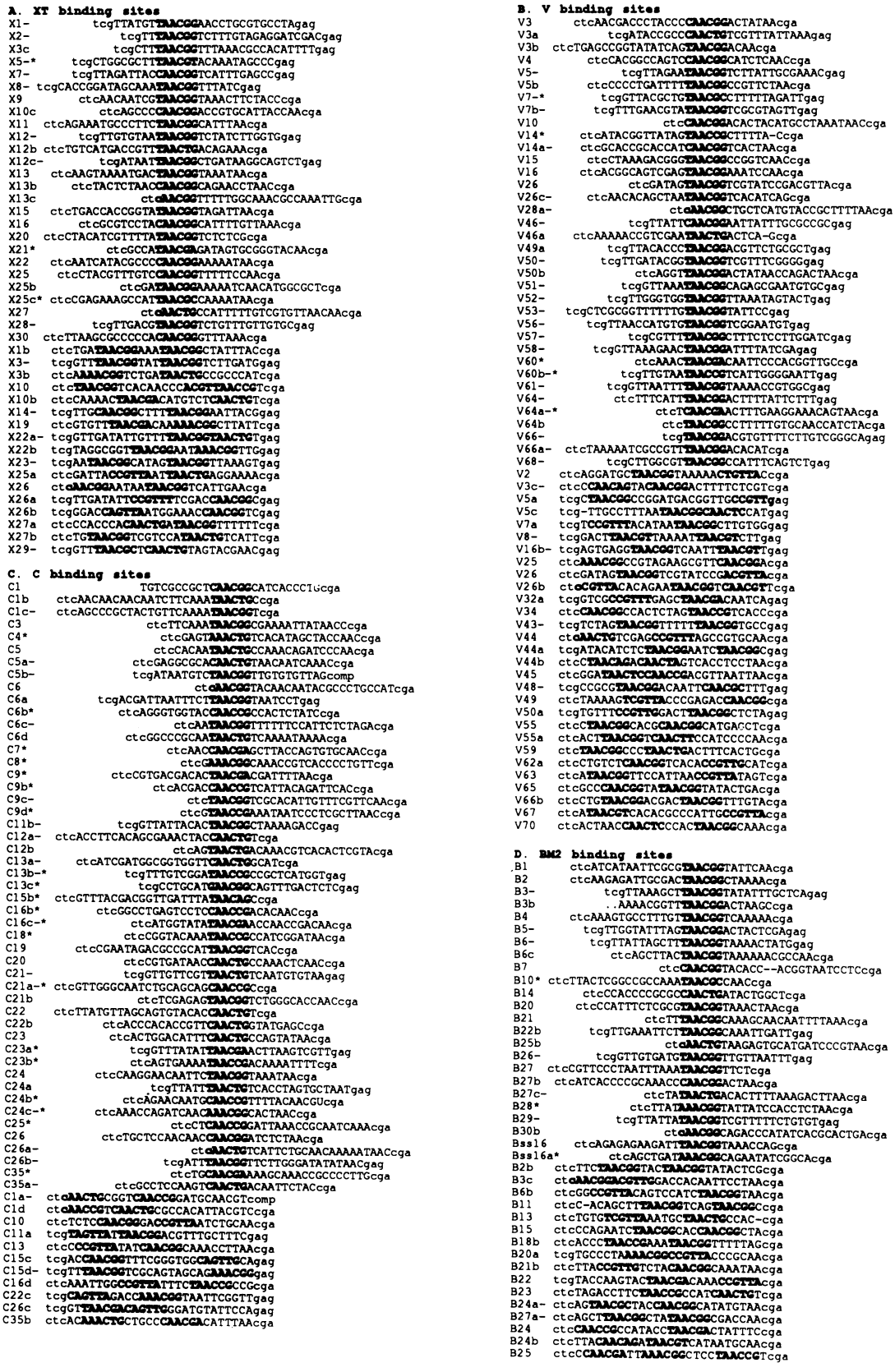

Figure 3. Sequences of cloned selected round 4 oligonucleotides. Minus signs indicate reverse orientation to original sequence reading. Starred sequences contain a 5/6 match to the Myb core consensus (shown in bold type); the remainder contain a 6/6 match. Single site sequences are shown aligned with respect to the Myb core; double or triple site sequences are presented unaligned. Nucleotides in lower case are the ends of the fixed flanking regions. Comp: ambiguous sequence caused by compression artefact on gel.

Table 1. Analysis of selected sequences. Base preferences for single sites, double sites, and all sites together are shown for each of the four protein sources. The bases immediately 5′ of the myb core hexamer are numbered +1 to +5; the core is unnumbered. Base preferences of all sites are shown as simple scores, and also as percentages.

A. XT binding sites

Single sites									
T	17		2	1	12	6	12	8	16
C	7		-	1	6	9	2	2	1
G	-		24	23	2	-	3	3	1
A	-		-	1	6	11	9	13	8
Total	24	26	26	26	26	26	26	26	26
Position		AAC			+1	+2	+3	+4	+5
Double sites									
T	24		8	1	18	8	12	13	10
C	6		1	1	8	10	1	4	5
G	-		26	31	2	2	7	2	2
A	4		-	2	7	12	11	11	11
Total	34	36	36	36	36	32	31	30	28
Position		AAC			+1	+2	+3	+4	+5
All sites									
T	41		10	2	30	14	24	21	26
C	13		1	2	14	19	3	6	6
G	-		50	54	4	2	10	5	3
A	4		-	3	13	23	20	24	19
Total	58	61	61	61	61	58	57	56	54
Position		AAC			+1	+2	+3	+4	+5
All sites (%)									
T	70.7		16.4	3.3	49.2	24.1	42.1	37.5	48.2
C	22.4		1.6	3.3	23.0	32.8	5.3	10.7	11.1
G	0		82.0	88.5	6.6	3.5	17.5	8.9	5.6
A	6.9		0	4.9	21.3	39.7	35.1	42.9	35.2
Position		AAC			+1	+2	+3	+4	+5

B. V binding sites

Single sites									
T	27		2	-	12	6	14	19	20
C	7		2	1	11	23	-	8	3
G	-		32	33	-	-	10	3	3
A	1		-	2	13	7	12	6	9
Total	35	36	36	36	36	36	36	36	35
Position		AAC			+1	+2	+3	+4	+5
Double sites									
T	36		7	7	22	12	15	13	16
C	17		1	4	22	19	9	12	13
G	-		48	44	1	4	7	7	2
A	4		3	3	10	17	21	17	15
Total	57	59	59	58	57	52	52	49	46
Position		AAC			+1	+2	+3	+4	+5
All sites									
T	63		9	7	34	18	29	32	36
C	24		3	5	35	42	9	20	16
G	-		80	77	1	4	17	10	5
A	5		3	5	23	24	33	23	24
Total	92	95	95	94	93	88	88	85	81
Position		AAC			+1	+2	+3	+4	+5
All sites (%)									
T	68.5		9.5	7.5	36.6	20.5	33.0	37.7	44.4
C	26.1		3.2	5.3	37.6	47.7	10.2	23.5	19.8
G	0		84.2	81.9	1.1	4.6	19.3	11.8	6.2
A	5.4		3.2	5.3	24.7	27.3	37.5	27.1	29.6
Position		AAC			+1	+2	+3	+4	+5

C. C binding sites

Single sites									
T	25		16	-	20	7	7	18	12
C	18		10	-	16	23	5	9	6
G	1		22	44	4	2	4	3	3
A	3		1	5	9	11	27	13	20
Total	47	49	49	49	49	43	43	43	41
Position		AAC			+1	+2	+3	+4	+5
Double sites									
T	9		7	-	6	2	3	5	5
C	8		3	-	9	8	3	7	7
G	-		12	19	3	2	6	1	3
A	3		-	3	1	6	6	4	2
Total	20	22	22	22	19	18	18	17	17
Position		AAC			+1	+2	+3	+4	+5
All sites									
T	34		23	-	26	9	10	23	17
C	26		13	-	25	31	8	16	13
G	1		34	63	7	4	10	4	6
A	6		1	8	10	17	33	17	22
Total	67	71	71	71	68	61	61	60	58
Position		AAC			+1	+2	+3	+4	+5
All sites (%)									
T	50.8		32.4	0	38.2	14.8	16.4	38.3	29.3
C	38.8		18.3	0	36.8	50.8	13.1	26.7	22.4
G	1.5		47.9	88.7	10.3	6.6	16.4	6.7	10.4
A	9.0		1.4	11.3	14.7	27.9	54.1	28.3	37.9
Position		AAC			+1	+2	+3	+4	+5

D. BM2 binding sites

Single sites									
T	17		3	-	13	4	6	5	6
C	3		-	1	6	6	2	3	6
G	-		21	23	-	-	4	1	1
A	2		-	-	5	14	12	15	9
Total	22	24	24	24	24	24	24	24	22
Position		AAC			+1	+2	+3	+4	+5
Double sites									
T	21		2	2	11	8	5	11	11
C	9		4	1	19	9	5	5	5
G	-		26	26	-	1	4	2	-
A	2		1	4	3	12	15	8	8
Total	32	33	33	33	33	30	29	26	24
Position		AAC			+1	+2	+3	+4	+5
All sites									
T	38		5	2	24	12	11	16	17
C	12		4	2	25	15	7	8	11
G	-		47	49	-	1	8	3	1
A	4		1	4	8	26	27	23	17
Total	54	57	57	57	57	54	53	50	46
Position		AAC			+1	+2	+3	+4	+5
All sites (%)									
T	70.4		8.8	3.5	42.1	22.2	20.8	32.0	37.0
C	22.2		7.0	3.5	43.9	27.8	13.2	16.0	23.9
G	0		82.5	86.0	0	1.9	15.1	6.0	2.2
A	7.4		1.8	7.0	14.0	48.2	50.9	46.0	37.0
Position		AAC			+1	+2	+3	+4	+5

completely to abolish binding under bandshift conditions (probe 2K, lane 4), indicating that differences must exist in the binding affinities of the four possible myb core sequences (see Discussion). Oligonucleotides 3K and 4K, which contain G residues at positions +1, +2, +4 and +5 (3K) or positions +1 and +2 (4K) relative to the Myb hexamer core, failed to bind to any of the three proteins (lanes 5 and 6); oligonucleotides 5K and 6K, which contain a G residue at either +1 or +2, appear to bind either very weakly (lane 7) or not at all (lane 8); probe 7K, with an A residue at position +1, is functional, but less efficient than 9M1 (lane 9), and probe 8K, designed as an optimal site for c-Myb, binds approximately as well as probe 7K (lane 10). It therefore appears that the extended consensus derived using the site selection assay is correct, with respect to the exclusion of G residues from the 3′ side of the myb core site.

Table 2. Consensus Myb binding sites derived from analysis of selected sequences. For each of the four protein sources, the consensus sequence for positions +1 to +5 to the 5′ side of the myb core hexamer was derived as follows: >N = base occurred in >50% of cases; N/N = either base occurred in >75% of cases; notG = G occurred in <6.5% of cases, and lowG = G occurred in ~10% of cases. The general consensus was derived similarly.

XT	T/C	AAC	G/T	G	>T notG	notG	T/A	T/A	T/A notG
V	T/C	AAC	G	G	notG	>C notG	T/A	lowG	>T notG
C	T/C	AAC	G/T/C	G	T/C	>C notG	>A	notG	lowG
BM2	T/C	AAC	G	G	T/C notG	>A notG	>A	T/A notG	notG
Together	Y	AAC	G/T(C)	G	notG	notG	-	lowG	notG
Position	1	234	5	6	+1	+2	+3	+4	+5

DISCUSSION

A site selection assay has been employed to show that the chicken c-Myb protein and the AMV v-Myb oncoprotein share a preferred common binding site, with consensus YAACKGHH. In the case of c-Myb, the consensus can reasonably be extended to include a C residue at position 5, as sites containing this residue were detected at greater frequency (18%) than when v-Myb proteins were used (4%). Additionally, slightly different binding preferences outside the originally defined Myb core sequence were detected for each of the four proteins. It should be stressed that none of the preferences were mutually exclusive, and may be due to the conditions of the assay. During initial rounds of site selection, the selecting protein will be in excess over its cognate binding sites, and will therefore recognise a complete spectrum of both low and high affinity sites; however, as cognate

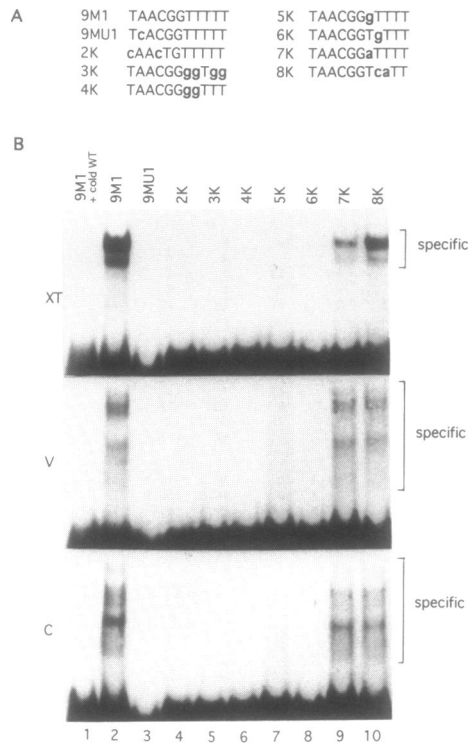

Figure 4. Verification of an extended Myb consensus sequence. A. Oligonucleotides used for gel mobility shift assays. Mutations with respect to the 9M1 parent oligonucleotide are shown in bold type. B. Gel mobility shift assays with XTMyb, v-Myb and c-Myb. Lane 1 contains a 100-fold excess of cold 9M1 oligonucleotide. Reactions loaded in all other lanes did not contain oligonucleotide competitors. Note that a slow-migrating non-specific complex is visible in some lanes; this is not a Myb-specific band as it is detected in conditions where Myb cannot bind (lane 3, probe 9MU1).

sites are amplified during succeeding rounds of selection, the number of binding sites will eventually exceed the amount of protein, and so there will be preferential binding to higher rather than low affinity sites, meaning that only a subset of all possible binding sites will be represented. The initial spectrum of sites selected and the point at which the change to high versus low affinity selection occurs is dependent on the binding affinity and concentration of the selecting protein. As both these parameters differed for the four proteins used here, differences in base preference might be expected. Consensus sequences derived using v-Myb protein from BM2 nuclear extract and the in vitro translated proteins may also have differed due to dissimilarities in protein modification, such as phosphorylation, affecting binding affinity.

Enumeration of the frequency of occurrence of all four possible 'perfect' core hexamers in the selected sequences showed that TAACGG was selected at least 50% of the time by the v-Myb derivatives (XT: 27/49; V: 46/67; BM2: 27/40), and 38% of the time by c-Myb. TAACTG and CAACTG were the least popular permutations for the v-Myb derivatives (XT: 6/49 and 4/49 respectively; V: 2/67 and 2/67 respectively; BM2 : 2/40 and 2/40 respectively) but occurred slightly more often in the c-Myb selection (7/42 and 12/42 respectively), as would be expected from the less stringent base preferences of c-Myb at positions 1 and 5 of the Myb core site (see Results). This difference between the v-Myb derivatives and c-Myb may again be a consequence of the site selection assay; as c-Myb binds DNA least well of the four test proteins (15), lower affinity sites may still be present in the round 4 population of oligonucleotides, for the reason discussed above. However, the lower frequency of occurrence of these two binding sites in the XT, V and BM2 site selection assays may be a reflection of their weaker affinity for v-Myb protein; c-Myb may not differentiate quite so much between the four possible Myb cores. The *in vivo* consequences of this difference with respect to transformation by v-Myb are hard to predict. As v-Myb has a higher affinity for DNA than does c-Myb (20), the elevated amounts of v-Myb necessary for transformation (21) might be expected to bind most or all of those sequences recognised by c-Myb. The increased binding specificity of v-Myb over c-Myb observed here does not support the theory that AMV v-Myb binds an additional spectrum of transformation-specific sites.

A noticeable feature of the selected oligonucleotides was that many contained two putative Myb binding sites. A second Myb core hexamer would be expected to occur by chance in a 26-mer about every 36 molecules, assuming the maximum possible spacing between the two motifs; in practice, spacing was usually less than maximum in the double site oligonucleotides observed, so the probability is actually less than this. As between 1 in 6 (for c-Myb) and 1 in 3 (for BM2) of oligonucleotides carried double sites, it must be assumed that this is of significance, and presumably means that Myb proteins have a greater binding affinity for double than single sites. This predilection for double sites has also been noted in a similar study using the murine c-Myb protein (19), but the basis for the phenomenon is unclear; possibly, some cooperativity of binding is occurring. It is of interest to note that in the promyelocyte-specific *mim*-1 gene, which is the only clearly defined gene shown to be directly regulated *in vivo* by Myb proteins (17), trans-activation occurs via three closely spaced binding sites in the promoter region.

The base preferences indicated by site selection for the Myb proteins were broadly confirmed by gel mobility shift assays using specific oligonucleotide probes. Whilst any mutation of the binding site 9M1 appeared to decrease binding affinity, the presence of G residues at positions +1 and +2 relative to the Myb core site caused the most dramatic reduction of binding for all the Myb proteins in gel mobility assays, as predicted from the site selection data. Inspection of selected oligonucleotides showed that only 2/284 putative myb core hexamers were followed by the sequence GG, whose presence completely abolished binding in the gel mobility assays, and only 19/284 had a G at either +1 or +2, which reduced binding to almost undetectable levels. Interestingly, the dinucleotide GG occurs in only 1/284 sites at positions +4 and +5, and indeed only 8/284 times between positions +1 to +5, making it seem likely that it has deleterious effects on binding in positions other than +1 and +2. Further experiments are required to test this possibility.

An oligonucleotide in which the high affinity Myb core TAACGG was mutated to CAACTG failed to complex significantly with any of the chicken Myb proteins in gel mobility shift assays. Bona fide selected binding sites unable to form complexes in gel mobility shift assays have been described previously (19). Presumably, the sensitivity of the gel mobility shift assay is lower than that of site selection with regard to the detection of all possible complexes; the PCR amplification step in site selection ensures that even complexes present in small quantities can be scored. Bearing this in mind, the gel mobility shift results provide confirmation that the CAACTG core is a lower affinity binding site for the Myb proteins, as predicted by site selection.

In a similar study using a C-terminally truncated murine c-Myb protein (19), an extended Myb consensus sequence of YAA-$C^{T/}C_{/G}$GYCR was derived from comparison of 51 selected oligonucleotides. This sequence is sufficiently similar to the chicken c-Myb consensus derived here of YAA$C^{T/}(C)_{/G}$GYCA to make it seem likely that the murine and chicken proteins have almost identical binding properties. As there is only one amino acid difference between the DNA binding domains of the murine and chicken proto-oncogenes, that they share a consensus sequence is not unexpected.

In conclusion, the three amino acid changes in the DNA binding domain of the AMV v-Myb oncoprotein relative to that of E26 and c-Myb do not cause any major changes in binding specificity, as assayed by site selection. Comparison of the selected binding sites of AMV v-Myb and c-Myb indicates that the preferred v-Myb core consensus is YAAC^{T} $/_{G}$G, whereas that of c-Myb is YAA$C^{T/}(C)_{/G}$G. The apparent relaxation in the specificity of c-Myb binding is probably due to its reduced affinity for DNA relative to v-Myb; the sequence YAACCG was observed in site selections with the three v-Myb derivatives, but at much lower frequency. The observed differences in the transformation phenotype caused by mutation of any of the three residues in the DNA binding domain are therefore unlikely to be due to changes in binding specificity. Other potential mechanisms, such as differential recognition of accessory factors, or changes in binding affinity, must be invoked instead.

ACKNOWLEDGEMENTS

I thank Pratipa Badiani for skilled technical assistance, and Mary Collins for helpful comments on the manuscript. This work was supported by the Cancer Research Campaign.

REFERENCES

1. Weston, K. and Bishop, J. M. (1989) *Cell*, **58**, 85–93.
2. Klempnauer, K.-H. and Sippel, A. E. (1987) *EMBO J.*, **6**, 2719–2725.
3. Peters, C. W. B., Sippel, A. E., Vingron, M. and Klempnauer, K.-H. (1987) *EMBO J.*, **6**, 3085–3090.
4. Frampton, J., Leutz, A., Gibson, T. and Graf, T. (1989) *Nature*, **342**, 134.
5. Howe, K. M., Reakes, C. F. L. and Watson, R. J. (1990) *EMBO J.*, **9**, 161–169.
6. Oehler, T., Arnold, H., Biedenkapp, H. and Klempnauer, K.-H. (1990) *Nucl. Acids Res.*, **18**, 1703–1710.
7. Introna, M., Golay, J., Frampton, J., Nakano, T., Ness, S.A. and Graf, T. (1990) *Cell*, **63**, 1287–1297.
8. Norman, C., Runswick, M., Pollock, R. and Treisman, R. (1988) *Cell*, **55**, 989–1003.
9. Klempnauer, K.-H., Gonda, T. J. and Bishop, J. M. (1982) *Cell*, **31**, 453–463.
10. Melton, D. A., Krieg, P.A., Rebagliati, M., Maniatis, T., Zinn, K. and Green, M.R. (1984) *Nucl. Acids Res.*, **12**, 7035–7056.
11. Jackson, R. J. and Hunt, R. T. (1983) *Meth. in Enzymol.*, **96**, 50–74.
12. Evan, G. I., Lewis, G. K. and Bishop, J. M. (1984) *Mol. Cell. Biol.*, **4**, 2843–2850.
13. Dignam, J. D., Lebovitz, R. M. and Roeder, R. G. (1983) *Nucl. Acids Res.*, **11**, 1475–1489.
14. Pollock, R. and Treisman, R. (1990) *Nucl. Acids. Res.*, **18**, 6197–6204.
15. Ramsay, R. G., Ishii, S. and Gonda, T. J. (1991) *Oncogene*, **6**, 1875–1879.
16. Klempnauer, K.-H., Symonds, G., Evan, G. I. and Bishop, J. M. (1984) *Cell*, **37**, 537–547.
17. Ness, S. A., Marknell, A. and Graf, T. (1989) *Cell*, **59**, 1115–1125.
18. Biedenkapp, H., Borgmeyer, U., Sippel, A. E. and Klempnauer, K.-H. (1988) *Nature*, **335**, 835–837.
19. Howe, K. M. and Watson, R. J. (1991) *Nucl. Acids Res.*, **19**, 3913–3919.
20. Luscher, B., Christenson, E., Litchfield, D. W., Krebs, E. G. and Eisenman, R. N. (1990) *Nature*, **344**, 517–522.
21. Schirm, S., Moscovici, G. and Bishop, J. M. (1990) *J. Virol.*, **64**, 767–773.