MOLECULAR AND CELLULAR BIOLOGY, Aug. 1991, p. 4104–4110
0270-7306/91/084104-07$02.00/0

Vol. 11, No. 8

Cyclic Amplification and Selection of Targets (CASTing) for the Myogenin Consensus Binding Site

WOODRING E. WRIGHT,* MICHAEL BINDER, AND WALTER FUNK

Department of Cell Biology and Neuroscience, University of Texas Southwestern Medical School, 5323 Harry Hines Boulevard, Dallas, Texas 75235

Received 26 December 1990/Accepted 13 May 1991

The consensus binding site for the muscle regulatory factor myogenin was determined from an unbiased set of degenerate oligonucleotides using CASTing (cyclic amplification and selection of targets). Stretches of totally random sequence flanked by polymerase chain reaction priming sequences were mixed with purified myogenin or myotube nuclear extracts, DNA-protein complexes were immunoprecipitated with an antimyogenin antibody, and the DNA was amplified by polymerase chain reaction. Specific binding was obtained after four to six cycles of CASTing. The population of selected binding sites was then cloned, and a consensus was determined from sequencing individual isolates. Starting from a pool with 14 random bases, purified myogenin yielded a consensus binding site of AACAG[T/C]TGTT, while nuclear extracts retrieved the sequence TTGCACCTGTTNNTT from a pool containing 35 random bases. The latter sequence is consistent with that predicted from combining an E12/E47 half-site (N[not T]CAC) with the purified myogenin half-site ([T/C]TGTT). The presence of paired E boxes in many of the sequences isolated following CASTing with nuclear extracts proves that myogenin can bind cooperatively with other E-box-binding factors.

Myogenin is a basic-helix-loop-helix (bHLH) protein (18) that is a member of a family of regulatory factors involved in the determination and differentiation of skeletal muscle (3, 4, 9, 10, 17, 22, 32). The muscle bHLH proteins can bind DNA sequences such as the MEF-1 site (8, 15) within the creatine kinase enhancer. Cotransfections of these muscle regulatory factors along with reporter genes containing the MEF-1 or similar sites into nonmuscle cells suggests that these proteins function directly or indirectly as transcription factors (3, 4, 10, 15, 16, 21, 30, 31, 33), and at least one such transcriptional activating domain has been identified (5).

Most of the actions of the muscle bHLH proteins are thought to occur as heterodimers with a ubiquitous bHLH protein of the E2A gene family (3, 6, 15, 19). These actions may be inhibited by Id (inhibitor of differentiation), an HLH protein that lacks the basic domain and is thus unable to bind DNA. It is hypothesized that Id forms inactive complexes with the E2A proteins and sequesters them so that they are not available for heterodimerization with the muscle bHLH factors (1). Those muscle bHLH proteins found in proliferating myoblasts in culture (MyoD and Myf-5) would thus be inactive until the cells were stimulated to differentiate, at which time the decline in Id levels would release the E2A proteins for heterodimer formation with the muscle bHLH proteins.

One mechanism of maintaining the commitment to the myogenic lineage is thought to involve an autoregulatory circuit in which MyoD stimulates its own expression (28). Under this model, autoregulation would occur in myoblasts, and thus must function in the presence of Id and the relative absence of free E2A subunits. This implies that some muscle bHLH proteins may be active either as homodimers or in complexes with proteins other than E2A.

To better understand the function of myogenin and its relation to the other skeletal bHLH proteins, we developed a method that would permit us to determine the myogenin consensus binding site from a pool of totally random oligonucleotides. This technique (CASTing [cyclic amplification and selection of targets]) has been used to determine the consensus binding site for myogenin homomers starting from a totally unbiased 270 million-fold degenerate oligonucleotide and can be applied as a general method for the determination of DNA binding sequences for any protein with affinity for nucleic acids. The consensus for myogenin homomers was then compared with the sequences obtained when the myogenin complexes present in myotube nuclear extracts were used to retrieve their binding sites from a pool of 10^{14} random sequences. The results indicate that myogenin homodimers bind a sequence closely related to the MEF-1 binding site and that the affinity of myogenin homomers for its optimal binding site is approximately the same as that for myogenin/E12 heterodimers. The binding sites obtained by using nuclear complexes are consistent with the combination of two half-sites, composed of half of the myogenin consensus site described here and the E12/E47 half-site recently reported and determined by using a closely related technique (2).

MATERIALS AND METHODS

An 84-mer was synthesized beginning with the M13 forward primer (GACGTTGTAAAACGACGGCCAGT), followed by a stretch of 14 totally random bases flanked by restriction sites (*Hin*dIII-*Pst*I-NNNNNNNNNNNNNN-*Bam*HI-*Eco*RI), and ending with sequences complementary to the M13 reverse primer (AATCATGGTCATAGCTGTT TCCT). A premixed solution of all four phosphoramidites was used during the synthesis of the random segment to ensure an equal incorporation of all four bases at each position. The presence of double restriction sites at each end permits two libraries of sequences to be cloned by using different pairs of enzymes, to eliminate the possibility that the consensus binding site contained one of the restriction sequences. The oligonucleotide was made double stranded by extension of 15 μg for 30 min at 72°C with *Taq* DNA

* Corresponding author.

polymerase in a standard reaction mixture (100 μl containing 10 mM Tris [pH 8.3], 50 mM KCl, 1.5 mM MgCl, 0.001% gelatin, and 200 μM deoxynucleoside triphosphate) with the M13 reverse primer as the sole polymerase chain reaction (PCR) primer.

A 5-μg sample of this double-stranded DNA was mixed with 0.2 μg of a purified myogenin fusion protein (fused to glutathione *S*-transferase [GS] [25]) in 20 μl of buffer (20 mM *N*-2-hydroxyethylpiperazine-*N'*-2-ethanesulfonic acid [HEPES; pH 7.6], 100 mM KCl, 0.2 mM EDTA, 20% glycerol, 10 mg of bovine serum albumin [BSA] per ml). After a 20-min incubation at room temperature, 10 μl of antimyogenin antibody-coated magnetic beads (30 mg of beads per ml, saturated by incubating goat anti-mouse-coated beads [Dynal Inc., Great Neck, N.Y.] with a threefold predicted excess of hybridoma culture supernatant) was added, and the mixture was agitated for 1 h at room temperature. Then 500 μl of isotonic saline containing 0.5% Nonidet P-40 and 0.1% BSA was added, and the bead/myogenin/DNA complexes were recovered by holding a 12-lb (ca. 5.4-kg) magnet (obtained from a local hardware store) against the side of the microfuge tube for 1 min before withdrawing the supernatant. Following three washes with the same solution and a final 20-μl buffer-exchange wash with PCR buffer, the beads were resuspended in 40 μl of a PCR reaction mixture containing 15 pmol of the M13 forward and reverse primers per 100 μl. A 10-μl portion of this mixture was reserved as an aliquot of the material present at each stage, and the remaining 30 μl was subjected to nine cycles of PCR (5 min at 100°C to denature the myogenin and antibodies, followed by the addition of *Taq* polymerase and eight thermal cycles of 94°C for 1 min, 65°C for 1 min, and 72°C for 1 min, after which the primer concentration was increased to 45 pmol/100 μl [see below] and a final cycle with an extension time of 10 min was performed). A 10-μl sample of the amplified DNA was used to initiate the next CASTing cycle; thus, a total of one-fourth of the material recovered in each cycle was carried forward to the next cycle.

It is important to limit the number of PCR cycles at each stage, since we found (data not shown) that the size of the amplified DNA became extremely large if excess cycles were used. This presumably resulted from artifacts introduced by futile cycles occurring after plateau was reached. An additional reason for avoiding plateau is that a polydisperse population of molecules is being amplified that has identical ends with a variable 14-bp region in the center. At plateau, the reassociation of different strands would be expected to produce heteroduplexes containing single-stranded centers, which would not serve as double-stranded targets for the next CASTing cycle. Since increasing the amount of primers can increase the plateau level of DNA, extra primers were added for the final cycle to ensure that the maximum fraction of sequences represented true double-stranded DNA. The time required for a CASTing cycle is about 2.5 hours; thus, a complete series of six cycles can be accomplished in 2 to 3 days.

Nuclear extracts were prepared as described by Lassar et al. (15a). For use, 2 μl of extract (in 0.5 M NaCl–20 mM HEPES [pH 7.5]–1.5 mM $MgCl_2$–10 mM dithiothreitol–1 mM phenylmethylsulfonyl fluoride–10 μg of pepstatin per ml–10 μg of leupeptin per ml–100 μg of aproteinin per ml–0.1% Triton X-100–20% glycerol) was diluted with 8 μl of the same buffer but lacking detergent and containing only 10 mM NaCl in order to reduce the salt concentration to about 100 mM. The random oligonucleotide used when CASTing with myotube nuclear extracts contained a central stretch of 35 rather than 14 random bases, for a total length of 105 bp. Since a substantial fraction of this oligonucleotide bound nonspecifically to the nuclear extracts, 10 μg of sonicated salmon sperm DNA was included as a competitor during the incubations. This reduced the actual amount of immunoprecipitated specific DNA/protein complexes far below what was obtained with use of the fusion protein. At each CASTing cycle, the recovered DNA was PCR amplified and aliquots were removed after 10, 15, and 20 thermal cycles. The aliquot from which DNA could first be faintly detected by ethidium bromide staining on agarose gels was used for the next CASTing cycle. As was the case with use of purified myogenin, overamplification of the DNA resulted in the production of very large DNA by the fourth or fifth cycle of CASTing.

Monoclonal antimyogenin antibodies were produced by injecting GS-myogenin in RIBI adjuvant (RIBI Immunochem Research, Inc., Hamilton, Mont.) into the hind footpads of BALB/c mice every 3 days for six injections (7). Popliteal lymph nodes were then excised, and the lymphocytes were fused to NS-1 myeloma cells. The monoclonal antimyogenin antibody used in these experiments, F5D, shifts the mobility of myogenin/DNA complexes on EMSA (electrophoretic mobility shift assays) and thus does not disrupt the protein-DNA interaction. It does not recognize either MyoD or Myf-5 (data not shown).

RESULTS AND DISCUSSION

The myogenin consensus binding site was determined by using CASTing. We synthesized an 84-base oligonucleotide containing a random set of 14 bases (making this a $4^{14} = 270$ million-fold degenerate oligonucleotide) flanked by restriction sites and PCR priming sequences. After synthesis of the complementary strand by priming DNA synthesis with the 3' PCR primer, the double-stranded oligomer (0.1 nmol, containing 10^5 copies of each of the random species) was mixed with a bacterially expressed myogenin fusion protein to permit the formation of myogenin/DNA complexes and then immunoprecipitated with magnetic beads coated with an antimyogenin monoclonal antibody. The myogenin and antibodies were denatured by boiling, the released DNA was reamplified and mixed with fresh myogenin, and the process was begun again. The series of steps of incubation, immunoprecipitation, and reamplification will be referred to as one cycle of CASTing.

Figure 1 demonstrates the progressive enrichment for myogenin binding sites that was obtained with this approach. During the first two cycles, the fraction of specific targets within the population was so small that no association with myogenin could be observed by using EMSA. However, a small signal was obtained during the third cycle, which became a major fraction of the total in the fourth through sixth cycles. The presence of two distinct bands in Fig. 1 reflects the presence of two different conformations of myogenin homodimers (unpublished data).

Figure 2 lists the sequences that were obtained after cloning the six-cycle-selected material. All of the sequences contained the E-box sequence of CANNTG (18), with most clones containing a CAGNTG core within a highly conserved 10-bp AACAG(C/T)TGTT sequence. This represents a distinctive subset of the consensus MEF-1 binding site (8), biased in favor of A/T and containing the frequent GT nonpalindromic center. The relative affinity of variations on the consensus sequence was examined by EMSA. Figure 3 demonstrates a virtually indistinguishable pattern with the

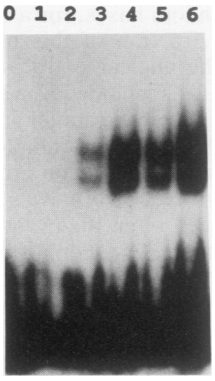

FIG. 1. Progressive enrichment for myogenin binding sites by using CASTing. The material obtained from each cycle of CASTing was radioactively labeled by including [^{32}P]dCTP during three cycles of PCR in the absence of cold dCTP and was subjected to EMSA in the presence of 100 ng of bacterially expressed myogenin. Specific binding could be detected as early as the third cycle and became strong by the fourth through sixth cycles.

different oligonucleotides, with no apparent differences between a palindromic E box (RC264; CAGCTG), the nonpalindromic E box (RC370; CAGTTG), a sequence containing an A instead of the highly conserved G at position 3 of the E box (RC315; CAATTG), or the MEF-1 site (CACCTG).

The muscle regulatory bHLH proteins are thought to function as heterodimers with E2A gene products such as E12 (3, 6, 15, 19). The affinity of myogenin homomers for the MEF-1 site was directly compared with that for myogenin/E12 heteromers to determine whether the affinity detected in Fig. 3 was comparable to that of the putative biologically active complex. Figure 4 demonstrates that with use of this class of targets, myogenin homodimers are competed for with kinetics very similar to those for myogenin/E12 heterodimers.

Sun and Baltimore have reported that a region adjacent to the basic domain in E12 prevents E12 homodimers from binding to DNA (27). Their E12 was bacterially expressed but was not a fusion protein. The GS-E12 fusion protein used in our experiments contained all of the coding sequences described by Murre et al. (19). It is possible that our GS-E12

RC87	AA	CAGCTG	TGTGA
RC92	AA	CAGCTG	TCA
RC126	GTA	CAGCTG	TAACA
RC182	GAA	CAGCTG	TTTTT
RC198	TTTGA	CAGCTG	CTA
RC205	AA	CAGCTG	TAACGC
RC207	GGGAA	CAGCTG	TTA
RC217	GTGT	CAGCTG	GACG
RC247	AA	CAGCTG	GTCTCA
RC257	GCTAG	CAGCTG	TCT
RC264	AA	CAGCTG	TTACTT
RC297	AA	CAGCTG	TCTCA
RC306	AA	CAGCTG	TTATAT
RC330	AA	CAGCTG	TCGTGC
RC331	AGT	CAGCTG	GTTTGA
RC338	AA	CAGCTG	TTACAT
RC342	GGAGT	CAGCTG	TCA
RC349	GTGCG	CAGCTG	TTC
RC364	TGTAA	CAGCTG	TTA

RC94	GAA	CAGTTG	GTTTC
RC127	GATGA	CAGTTG	TTA
RC132	AA	CAGTTG	TCTGA
RC149	A	CAGTTG	TCCTGGA
RC174	GTAA	CAGTTG	TTGG
RC177	AAA	CAGTTG	GTAAT
RC183	GCGT	CAGTTG	TCAC
RC197	CAA	CAGTTG	TTTCA
RC216	GTAA	CAGTTG	GCAC
RC242		CAGTTG	TTTTCTGCTCG
RC246	AA	CAGTTG	AGCTCA
RC253	AATCA	CAGTTG	ACC
RC265	ACGT	CAGTTG	TTTA
RC303	TAA	CAGTTG	CTTAT
RC319	AA	CAGTTG	TAGTGA
RC334	GGTGT	CAGTTG	TTA
RC351	TGAA	CAGTTG	TTA
RC356	GATTGA	CAGTTG	CT
RC367	GTTGA	CAGTTG	TAT
RC370	AAA	CAGTTG	TTTCT

RC90	GA	CATTTG	TTCGCT
RC302	ACGA	CACATG	TTAC
RC315	AAG	CAATTG	TTGC
RC354	GTGCG	CACGTG	TTC
RC357	CGT	CACATG	TCACT

G	8	8	7	13	4	-	-	39	1	-	44	6	2	4	5	5
A	2	4	5	25	32	-	-	1	2	-	-	2	5	18	3	7
T	2	7	11	1	7	-	44	1	22	44	-	33	26	13	10	6
C	-	1	5	3	-	44	-	3	19	-	-	3	11	8	12	5
				A	A	C	A	G	T/C	T	G	T	T			

MEF-1 consensus sequence $^{C}_{G}N^{GG}_{AA}CA^{CC}_{GG}TG^{CC}_{TT}N^{C}_{G}$

FIG. 2. CASTing of selected myogenin homomer binding sites. Sequences present after six cycles of CASTing were digested with either *Bam*HI-*Hin*dIII or *Eco*RI-*Pst*I and cloned into pBluescript II KS+ (Stratagene, La Jolla, Calif.). Individual colonies were picked and sequenced. Sequences are arranged according to their core E box into three groups: CAGCTG, CAGTTG, and CANNTG. The total distribution of nucleotides represents the sum of all three groups and is compared with the consensus MEF-1 binding site described by Buskin and Hauschka (8).

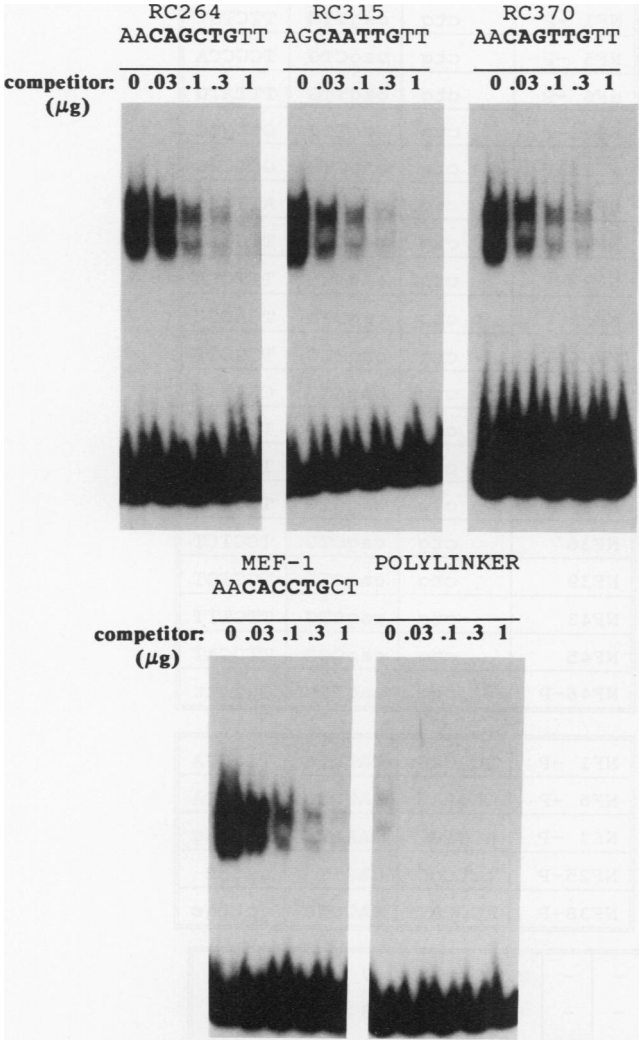

FIG. 3. Relative affinity of different selected sequences for myogenin. Individual cloned CASTing or control sequences were mixed with 50 ng of GS-myogenin and increasing amounts of competitor (sonicated salmon sperm DNA) and were then analyzed by EMSA. A 25-bp sequence containing a single MEF-1 binding site (kindly provided by Jean Buskin) was cloned into pBluescript II KS+ (Stratagene). Probe DNAs were labeled by PCR, using primers immediately flanking the *Bam*HI and *Hin*dIII cloning sites. The polylinker control represents the 24 bp between these sites in the parent vector. The 10-bp core sequences are listed above each set of competitions; the full sequences can be found in Fig. 2. Myogenin has approximately the same affinity to all of these representative binding sites.

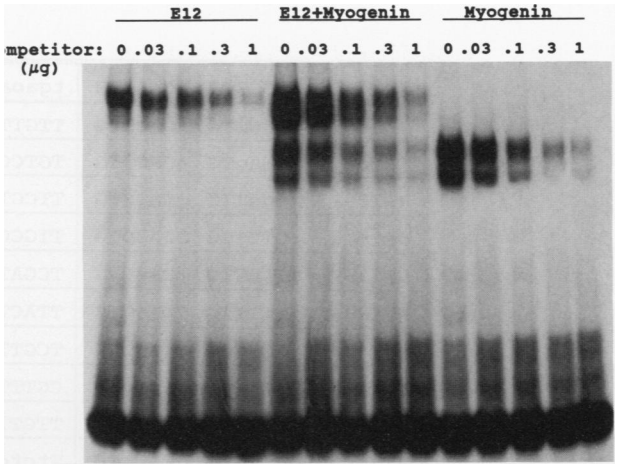

FIG. 4. Relative affinity of the MEF-1 site for E12 and myogenin homodimers versus E12/myogenin heterodimers. One nanogram of a 25-mer sequence containing the MEF-1 binding site was mixed with increasing amounts of sonicated salmon sperm DNA, incubated with 25 ng of GS-E12, 25 ng of GS-E12 preincubated for 1 h at 4°C with 50 ng of GS-myogenin, or 50 ng of GS-myogenin then analyzed by EMSA. One-half as much GS-E12 was used to reduce the intensity of the E12 homodimer band, which is not well resolved from the E12/myogenin heterodimer band in these gels. The relative affinity of myogenin homomers for the MEF-1 site is virtually indistinguishible from that of E12/myogenin heterodimers.

fusion protein is able to recognize the MEF-1 binding site (Figure 4) despite the presence of this inhibitory domain.

The consensus sequence for MyoD binding, determined by using a technique related to CASTing has recently been reported (2). The MyoD consensus sequence, (A/G)(A/t) CAGCTG(T/a)(T/C) (where lowercase letters represent a less utilized base), is virtually identical to the myogenin consensus shown in Fig. 2. The frequency of G and C at the distal positions of the myogenin consensus is about 25%, and since this is the frequency expected of a random base, we have not chosen to formally include these bases in our consensus sequence. Given the extraordinary similarity of the consensus binding sites, it is not surprising that the two muscle-specific bands in myotube nuclear extracts that bind to the MEF-1 site represent complexes between either MyoD or myogenin and the E2A gene product (15a). Preliminary information from several laboratories (personal communications) suggests that some sites exist in which myogenin binding or transactivation can be observed in the absence of MyoD binding or transactivation. If confirmed, these results might indicate that nucleotides outside the consensus sequence shown in Fig. 2 influence the relative binding of myogenin versus MyoD or that myogenin binding or activation is affected by the presence of additional proteins that might interact with specific bHLH factors.

Evidence for such interactions was sought by performing CASTing with myotube nuclear extracts, in which myogenin is complexed with its normal biological partner and which may contain other potentially interacting factors. Because these interactions might produce a much larger consensus sequence, an oligonucleotide containing a stretch of 35 random bases was used as the initial starting target. The 10 μg of this 105-mer used to initiate CASTing contained a small sampling of the total sequences present ($4^{23.5} = 10^{14}$ of the $4^{35} = 10^{21}$ possible sequences). Despite this extraordinary complexity, a very strong consensus sequence emerged that confirmed these predictions.

Figure 5 presents data from the 31 clones of cycle 6 CASTing DNA that we sequenced. All 31 sequences analyzed contained at least one E box, while 14 of the 31 contained paired E boxes (each of a pair of E boxes within a single clone is designated −P in Fig. 5). Nineteen of the clones included the *Pst*I cloning site within an E box (see below). Since definitive statements about which E box was the direct binding site for myogenin versus other bHLH factors could not be made for paired sites, those clones containing single sites were used for the initial groupings

A

NF3 -P	AGGGCG	CACCTG	tgacat
NF8	TGGTTG	CACCTG	TTGTTA
NF10	gcagTG	CACCTG	TGTCCT
NF11-P	tggTTG	CACCTG	TTCGTT
NF14	GTTTCG	CACCTG	TTGCCC
NF13-P	TATTTG	CACCTG	TCGATT
NF16-P	TTTTTG	CACCTG	TTACTC
NF18	GGAGAA	CACCTG	TCGTTT
NF20	TAGTTG	CACCTG	CGTTTT
NF24	CCGTTG	CACCTG	TTGTTT
NF25-P	GTTTTG	CACCTG	Ttctcg
NF26	AGGTTN	CACCTG	TCGTTT
NF27-P	AGTATN	CACCTG	TTNTtt
NF27-P	ggatc	cACCTG	CTAAaa
NF30-P	ccACGG	CACCTG	TTTACA
NF31	AGGTTA	CACCTG	TTCAAT
NF32-P	CTATTG	CACCTG	TTctta
NF35-P	GATGCA	CACCTG	TTACGT
NF38-P	ggtttg	CACCTG	TTCTTG
NF42	AGATTG	CACCTG	TTCA
NF46-P	gtgttG	CACCTG	TTGTTT

B

NF1 -P	ctg	cagTTG	TTCTGT
NF5 -P	ctg	cagCTG	TCGCCA
NF6 -P	ctg	cagTTG	TTTATG
NF11-P	ctg	cagTTG	Gttgca
NF13-P	ctg	cagCTG	GCGCGG
NF15-P	ctg	caaTTG	ATTGAC
NF16-P	ctg	cagGTG	TTGTAT
NF22-P	ctg	cagCTG	TTTGCA
NF28	ctg	cagCTG	TCACCC
NF30-P	ctg	cagTTG	TTGGCT
NF32-P	ctg	cagGTG	CGAtaa
NF33	ctg	cagCTG	TTGTTT
NF34	ctg	cagCTG	TTGTTT
NF35	ctg	cagCTG	TCGTCT
NF36	ctg	cagCTG	TCGTCT
NF39	ctg	cagCTG	TCGTGT
NF43	ctg	cagCTG	TTGTTT
NF45	ctg	cagCTG	TTCGAT
NF46-P	ctg	cagTTG	Ggtgtt

NF1 -P	ACACAG	CATATG	TTCCGA
NF6 -P	CGGAAC	CAGCTG	TTCCTA
NF3 -P	ACGTGC	CAAATG	tcacag
NF25-P	tctcgG	CAGTTg	gatcc
NF38-P	AGGCAT	CATCTG	gtttgc

C

G	4	7	6	3	1	14	-	-	-	-	-	21	-	2	7	1	1	1
A	5	3	4	1	1	3	-	21	-	-	-	-	-	-	3	5	1	2
T	4	4	6	12	13	-	-	-	-	-	21	-	18	14	3	8	10	10
C	2	1	-	1	3	-	20	-	21	21	-	-	2	3	4	4	3	2
				T	T	G	C	A	C	C	T	G	T	T	N	N	T	T

FIG. 5. Myogenin binding sites from CASTing with use of nuclear extracts. Each of a pair of E boxes within a single clone is designated −P. Constrained bases (within 6 bases of a second E box or contained within a cloning site flanking the stretch of 35 random bases) are shown in lowercase letters. (A) All **CACCTG**-containing sequences; (B) sequences in which the E box overlapped the *Pst*I cloning site; (C) All of the remaining E boxes. The total distributions presented at the bottom are based solely on the **CACCTG**-containing sequences.

presented in Fig. 5. Nine clones contained single E boxes that were not part of the *Pst*I site, and all nine of these contained the E-box sequence CACCTG. Since CACCTG must therefore be part of the myogenin binding site, we organized the sequences into three groups: those containing CACCTG, those constrained by their overlap with the *Pst*I site, and those in neither category. Initial inspection of these sequences indicated that the consensus extended as far as six bases beyond the E box. In many cases, the E boxes were located within six bases of a source of constraint (the cloning sites at the end of the 35-mer or another E box). Any nucleotide that is either part of a cloning site or constrained by its presence within six bases of another E box is shown in lowercase letters in Fig. 5. Constrained bases were excluded from the totals used for determining the consensus binding site shown at the bottom of Fig. 5.

The CACCTG group of myogenin binding sites has a very strong consensus sequence of TTG**CACCTG**TTNNTT. The E12/E47 half-site determined by Blackwell and Weintraub (N[not T]**CAC**) (2) is consistent with the 5′ half of this sequence, while the myogenin half-site shown in Fig. 5 (**CTGTT**) is fully contained within the 3′ half. Although consistent with the binding of myogenin/E2A product heterodimers, we will refer to this sequence as the myogenin heteromer site since the limited specificity of the E12/E47 consensus (essentially just CAC) prevents firm conclusions about the nature of the myogenin partner at that site. The NNTT sequence that extends beyond the myogenin consensus sequence of Fig. 5 may not have been observed in that experiment simply because of the difficulty of recognizing a 9-base half-site (and thus an 18-base total) in the limited window afforded by the 14 random bases used in the

CASTing with myogenin homomers. Alternatively, it may be due to a more extended region being recognized by myogenin heteromers or to the binding of an adjacent factor.

The *Pst*I recognition site (CTGCAG) is very similar to the myogenin heteromer consensus site. The TGCA sequence is present within the heteromer site, and the final G, present at the first N position of the E box (CANNTG), is consistent both with the myogenin consensus sequence and the MEF-1 consensus shown in Fig. 5. The nonrandom presence of the last five bases of the *Pst*I site thus introduces a 4^5 = 1,000-fold bias in favor of including this sequence. Because of this bias, we have not used the *Pst*I-constrained E boxes in calculating the myogenin heteromer consensus totals indicated at the bottom of Fig. 5. The presence of paired E boxes cannot be explained on the basis of this bias. The probability of retrieving an oligonucleotide with two independent (noninteracting) sites is only twice as great as that for a single site, and over six cycles of selection this would produce only a 32-fold enrichment for double sites. However, the consensus for the *Pst*I-constrained sites clearly extends seven to nine nucleotides into the random stretch, and thus the probability of finding these sequences along with a second site is at most one in 4^7 = 16,000. Their abundance is thus at least 500-fold greater than could be explained on the basis of the bias produced by the *Pst*I site. In addition, 3 of the 14 pairs did not utilize the *Pst*I site in either of the two E boxes.

Paired E boxes might theoretically be obtained if the affinity of the antibody for two myogenin complexes were greater than for a single complex. However, two factors essentially eliminate this possibility. First, since an excess of antibody was used, myogenin complexes are not in competition with each other for antibody binding. Second, the antibody is of sufficiently high affinity that bound myogenin is not released during the extremely gentle washing conditions used. A higher affinity of the antibody for paired versus single complexes thus would not provide a selective advantage for the retrieval of paired sites. Paired E boxes have been observed in a large number of muscle-specific enhancers and promoters (8). The retrieval of paired E boxes following CASTing with nuclear extracts provides unambiguous proof that the proteins that bind to these sites interact, either directly or via other factors, to stabilize the binding of the complexes to DNA. The ability of MyoD homomers to interact cooperatively with the paired sites in the creatine kinase enhancer (30) has already been demonstrated. This report establishes that myogenin can also participate in such cooperative interactions, although the partners in this process remain unresolved. A detailed analysis of pair spacing and additional regions (showing non-E-box consensus sequences due presumably to other associated nuclear factors) will be presented elsewhere.

Five E-box sequences fit into neither the CACCTG nor *Pst*I-constrained category. Because three of these sequences were highly constrained by nearby E boxes, there were insufficient numbers to analyze at the present time.

Tuerk and Gold (29) have found that differences in affinity of as little as twofold can affect the abundance of a sequence obtained by techniques such as CASTing. The dominance of the TTGCACCTGTTNTT sequence among the unconstrained sequences that we obtained following CASTing with myotube nuclear extracts thus does not imply that myogenin in nuclear extracts exists predominantly as heterodimers or that myogenin (either as homomers or heteromers) does not recognize other sequences. The results presented here indicate only that the observed consensus (in conjunction with the binding of adjacent proteins) provides the highest-affinity site present in the initial pool of 10^{14} sequences.

Previous work by several laboratories established the principle of using nucleotides with random bases at a limited number of specific sites in order to determine their effects on binding activity (12–14, 20, 23, 24). However, these approaches required that the specific nucleic acid binding site already be known, as well as being limited to randomizing a few bases within an overall fixed context. As this work was being completed, we became aware of several reports that exploited the concept of repetitive cycling to purify nucleic acid binding sites (2, 11, 14, 26, 29). These approaches reflect variations on one theme: the ability to amplify the material enriched by any procedure permits a modest level of purification to be repeated until very highly purified material is obtained. The name CASTing emphasizes this basic process and suggests the evocative image of using a known substance to fish for an unknown within a sea of random possibilities.

The recent study by Blackwell and Weintraub (2) used a related technique called SAAB to determine the MyoD and E12/E47 binding sites. In their approach, in vitro translations of cloned cDNAs were used to perform EMSA on oligonucleotides containing short random stretches, and then shifted bands were cut out of the gels and used for subsequent cycles. The MyoD consensus site that they obtained was virtually identical to the myogenin homomer site depicted in Fig. 5; thus, these sequences by themselves will not distinguish genomic targets regulated by one bHLH factor versus the other. Our use of antibodies to isolate protein/DNA complexes permitted us to apply CASTing to crude nuclear extracts as well as purified proteins. The results of using nuclear extracts were dramatic: not only were much larger direct consensus sequences obtained, but the footprints of proteins that interact with and increase the affinity of myogenin complexes for DNA were obtained. It is likely that the proteins with which myogenin, MyoD, MRF4, and Myf5 interact are different. CASTing with nuclear extracts and antibodies to the different bHLH factors should provide an opportunity to define different classes of recognition sequences for the higher-order complexes involving these proteins and thus provide insights into the different functions of these muscle regulatory factors.

ACKNOWLEDGMENTS

We acknowledge support from the National Institutes of Health and the Muscular Dystrophy Association of America.

REFERENCES

1. **Benezra, R., R. L. Davis, D. Lockshon, D. L. Turner, and H. Weintraub.** 1990. The protein Id: a negative regulator of helix-loop-helix DNA binding proteins. Cell **61**:49–59.
2. **Blackwell, T. K., and H. Weintraub.** 1990. Differences and similarities in DNA-binding preferences of myoD and E2A protein complexes revealed by binding site selection. Science **250**:1104–1110.
3. **Braun, T., E. Bober, B. Winter, N. Rosenthal, and H. H. Arnold.** 1990. *Myf-6*, a new member of the human gene family of myogenic determination factors: evidence for a gene cluster on chromosome 12. EMBO J. **9**:821–831.
4. **Braun, T., G. Buschhausen-Denker, E. Bober, E. Tannich, and H. Arnold.** 1989. A novel human muscle factor related to but distinct from MyoD1 induces myogenic conversion in 10T1/2 fibroblasts. EMBO J. **8**:701–709.
5. **Braun, T., B. Winter, E. Bober, and H. H. Arnold.** 1990. Transcriptional activation domain of the muscle-specific gene-regulatory protein myf5. Nature (London) **346**:663–665.

6. **Brennan, T. J., and E. N. Olson.** 1990. Myogenin resides in the nucleus and acquires high affinity for a conserved enhancer element on heterodimerization. Genes Dev. **4**:582–590.
7. **Brodsky, F. M.** 1985. Clathrin structure characterized with monoclonal antibodies. 1. Analysis of multiple antigenic sites. J. Cell Biol. **101**:2047–2054.
8. **Buskin, J. N., and S. D. Hauschka.** 1989. Identification of a myocyte nuclear factor which binds to the muscle-specific enhancer of the mouse creatine kinase gene. Mol. Cell. Biol. **9**:2627–2640.
9. **David, R. L., H. Weintraub, and A. B. Lassar.** 1987. Expression of a single transfected cDNA converts fibroblasts to myoblasts. Cell **51**:987–1000.
10. **Edmondson, D. G., and E. N. Olson.** 1989. A gene with homology to the *myc* similarity region of MyoD1 is expressed during myogenesis and is sufficient to activate the muscle differentiation program. Genes Dev. **3**:628–640.
11. **Ellington, A. D., and J. W. Szostak.** 1990. *In vitro* selection of RNA molecules that bind specific ligands. Nature (London) **346**:818–822.
12. **Horwitz, M. S. Z., and L. A. Loeb.** 1986. Promoters selected from random DNA sequences. Proc. Natl. Acad. Sci. USA **83**:7405–7409.
13. **Horwitz, M. S. Z., and L. A. Loeb.** 1988. DNA sequences of random origins as probes of *Escherichia coli* promoter architecture. J. Biol. Chem. **263**:14724–14731.
14. **Kinzler, K. W., and B. Vogelstein.** 1989. Whole genome PCR: application to the identification of sequences bound by gene regulatory proteins. Nucleic Acids Res. **17**:3645–3653.
15. **Lasser, A. B., J. N. Buskin, D. Lockshon, R. L. David, S. Apone, S. D. Hauschka, and H. Weintraub.** 1989. MyoD is a sequence-specific DNA binding protein requiring a *myc* homology to bind to the muscle creatine kinase enhancer. Cell **58**:823–831.

15a. **Lassar, A. B., et al.** Submitted for publication.

16. **Lin, Z.-Y., C. A. Deschesne, J. Eldridge, and B. M. Paterson.** 1989. An avian muscle factor related to MyoD1 activates muscle-specific promoters in nonmuscle cells of different germ-layer origin and in BrdU-treated myoblasts. Genes Dev. **3**:986–996.
17. **Miner, J. H., and B. Wold.** 1990. Herculin, a fourth member of the *MyoD* family of myogenic regulatory genes. Proc. Natl. Acad. Sci. USA **87**:1089–1093.
18. **Murre, C., P. S. McCaw, and D. Baltimore.** 1989. A new DNA binding and dimerization motif in immunoglobulin enhancer, *daughterless*, MyoD, and *myc* proteins. Cell **56**:777–783.
19. **Murre, C., P. S. McCaw, H. Vaessin, M. Caudy, L. Y. Jan, C. V. Cabrera, J. N. Buskin, S. D. Hauschka, A. B. Lassar, H. Weintraub, and D. Baltimore.** 1989. Interactions between heterologous helix-loop-helix proteins generate complexes that bind specifically to a common DNA sequence. Cell **58**:537–544.
20. **Oliphant, A. R., and K. Struhl.** 1987. The use of random-sequence oligonucleotides for determining consensus sequences. Methods Enzymol. **155**:568–582.
21. **Piette, J. L., J. L. Bessereau, M. Huchet, and J.-P. Chambon.** 1990. Two adjacent MyoD1 binding sites regulate expression of the acetylcholine receptor α-subunit gene. Nature (London) **345**:353–356.
22. **Rhodes, S. J., and S. F. Konieczny.** 1989. Identification of MRF4: a new member of the muscle regulatory factor gene family. Genes Dev. **3**:2050–2061.
23. **Schneider, T. D., and G. D. Stormo.** 1989. Excess information at bacteriophage T7 genomic promoters detected by a random cloning technique. Nucleic Acids Res. **17**:659–674.
24. **Singer, V. L., C. R. Wobbe, and K. Struhl.** 1990. A wide variety of DNA sequences can functionally replace a yeast TATA element for transcriptional activation. Genes Dev. **4**:636–645.
25. **Smith, D. B., and S. J. Johnson.** 1988. Single-step purification of polypeptides expressed in *Escherichia coli* as fusions with glutathione S-transferase. Gene **67**:31–40.
26. **Sompayrac, L., and K. J. Danna.** 1990. Method to identify genomic targets of DNA binding proteins. Proc. Natl. Acad. Sci. USA **87**:3274–3278.
27. **Sun, X.-H., and D. Baltimore.** 1991. An inhibitory domain of E12 transcription factor prevents DNA binding in E12 homodimers but not in E12 heterodimers. Cell **64**:459–470.
28. **Thayer, M. J., S. J. Tapscott, R. L. Davis, W. E. Wright, A. B. Lassar, and H. Weintraub.** 1989. Positive autoregulation of the myogenic determination gene MyoD1. Cell **58**:241–248.
29. **Tuerk, C., and L. Gold.** 1990. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. Science **249**:505–510.
30. **Weintraub, H., R. Davis, D. Lockshon, and A. Lassar.** 1990. MyoD binds cooperatively to two sites in a target enhancer sequence: occupancy of two sites is required for activation. Proc. Natl. Acad. Sci. USA **87**:5623–5627.
31. **Weintraub, H., S. J. Tapscott, R. L. Davis, M. J. Thayer, M. A. Adam, A. B. Lassar, and A. D. Miller.** 1989. Activation of muscle-specific genes in pigment, nerve, fat, liver, and fibroblast cell lines by forced expression of MyoD. Proc. Natl. Acad. Sci. USA **86**:5434–5438.
32. **Wright, W. E., D. A. Sassoon, and V. K. Lin.** 1989. Myogenin, a factor regulating myogenesis, has a domain homologous to MyoD. Cell **56**:607–617.
33. **Yutzey, K. E., S. J. Rhodes, and S. F. Konieczny.** 1990. Differential *trans* activation associated with the muscle regulatory factors MyoD1, myogenin, and MRF4. Mol. Cell. Biol. **10**:3934–3944.