Proc. Natl. Acad. Sci. USA
Vol. 89, pp. 9484–9488, October 1992
Developmental Biology

Cyclic amplification and selection of targets for multicomponent complexes: Myogenin interacts with factors recognizing binding sites for basic helix–loop–helix, nuclear factor 1, myocyte-specific enhancer-binding factor 2, and COMP1 factor

(DNA-binding factors/transcription complexes/muscle differentiation/myogenesis)

WALTER D. FUNK AND WOODRING E. WRIGHT*

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

Communicated by Harold Weintraub, June 11, 1992

ABSTRACT Myogenin is one of four muscle-specific basic helix–loop–helix regulatory factors involved in controlling myogenesis. We here describe various protein complexes that increase the affinity of myogenin for DNA. We mixed an oligonucleotide containing a degenerate center large enough to accommodate multiple binding sites with crude myotube nuclear extracts and used cyclic amplification and selection of targets with an antimyogenin monoclonal antibody to isolate protein–DNA complexes. Since each cycle of selection results in the enrichment for the sequences with the highest affinity, we isolated multicomponent sites in which myogenin binding was increased by its interaction with other DNA binding proteins. Myogenin interacts with members of the nuclear factor 1 family, the muscle-specific factor myocyte-specific enhancer-binding factor 2, and another factor, COMP1 (cooperates with myogenic proteins 1), that binds to the sequence TGATTGAC. Myogenin also exhibits cooperative binding with other proteins that recognize CANNTG motifs, and various constraints on spacing and orientation were observed. The application of this approach to other transcription factors should not only help identify the different functions of myogenin versus other members of the muscle basic helix–loop–helix regulatory family but also help define the general combinatorial mechanisms involved in eukaryotic gene regulation.

Myogenin (1, 2), MyoD (3), myf5 (4), and MRF4 (5–7) comprise a family of regulatory proteins that plays key roles in the determination and differentiation of skeletal muscle. These factors are part of a much larger family of basic helix–loop–helix (bHLH) proteins (8). Although the active species appears to be heterodimers, myogenin and MyoD can form tetrameric or larger complexes that fail to bind to known DNA sequences (9). All four of the skeletal muscle bHLH factors heterodimerize *in vitro* with members of the E2A and E2-2 bHLH family (10–13) and recognize the core sequence CANNTG.

We (14) and other laboratories (15–19) have developed techniques in which a protein selects a consensus binding site from a pool of oligonucleotides containing a degenerate core flanked by PCR primers. In these cyclic amplification and selection of targets (CASTing) techniques, DNA sequences are enriched based upon various criteria [immunoprecipitation (14, 15), binding to nitrocellulose (19), altered mobility in native gels (17), etc.]. The bound DNA is then PCR amplified and mixed with fresh protein, and the process is repeated. The ability to perform multiple cycles results in a simple enrichment step exponentially purifying a consensus site from a pool of at least 10^{14} random sequences (14).

Since all of the muscle bHLH proteins recognize the same core CANNTG binding sites with approximately equal affinity (20, 21), specificity should be a consequence of different protein–protein interactions. If those interactions involve other DNA binding proteins, the binding of a muscle bHLH heterodimer to its site might be stabilized by its participation in these multicomponent complexes. CASTing should provide an effective method for investigating these multicomponent interactions. Each cycle of CASTing results in the preferential isolation of those sequences with the highest affinity. If the binding of myogenin is stabilized by its interaction with other DNA binding proteins, those DNA sequences containing a binding site for myogenin and for a factor with which it interacts should "out-compete" DNA containing only a myogenin binding site. The DNA sequences recovered following CASTing should thus contain not only the myogenin site but also the sites for those factors for which cooperative binding exists. This hypothesis was tested by CASTing using crude myotube nuclear extracts (to provide myogenin and those factors with which it might interact), a synthetic oligonucleotide containing a sufficient number of internal random bases [35 base pairs (bp)] to accommodate multiple binding sites, and a monoclonal antimyogenin antibody to select for sequences based exclusively on the presence of myogenin in the complexes. We here report the confirmation of this hypothesis and the identification of cooperative interactions of myogenin with members of the nuclear factor 1 (NF1) and myocyte-specific enhancer-binding factor 2 (MEF2) families, the factor COMP1 (cooperates with myogenic proteins 1), and other likely bHLH factors.

MATERIALS AND METHODS

CASTing. CASTing was performed as described (14). Nuclear extract from C2C12 myotubes was mixed with 10 μg of a double-stranded oligonucleotide that contained 35 randomized bases centered between PCR priming sequences, and then protein–DNA complexes were immunoprecipitated with antimyogenin monoclonal antibodies. The DNA in myogenin complexes was PCR amplified, analyzing aliquots after 10, 15, and 20 cycles on 2% agarose gels. The DNA from the aliquot showing the first visible band (≈25 ng) was used to initiate the next CASTing cycle. After six cycles of selection, the DNA was cloned and sequenced. Electrophoretic mobility shift assay (EMSA) analysis was performed after amplifying the insert from each clone using end-labeled PCR primers flanking the cloning site. The data from three inde-

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: CASTing, cyclic amplification and selection of targets; bHLH, basic helix–loop–helix; EMSA, electrophoretic mobility shift assay; NF1, nuclear factor 1; MEF2, myocyte-specific enhancer-binding factor 2; COMP1, cooperates with myogenic proteins 1.

*To whom reprint requests should be addressed.

pendent sets of CASTing (a total of ≈60 sequences) were pooled in the present study.

The following abbreviations or conventions are used. Nonrandom bases due to restriction sites flanking the 35-bp degenerate center are indicated by lowercase letters; "N" indicates that nuclear extracts rather than purified protein were used; "F" and "C" indicate whether the antimyogenin monoclonal antibody F5D (whose epitope maps just carboxy-terminal to the bHLH region) or C12D (whose epitope maps just amino-terminal to the bHLH region) was used; " ' "indicates the use of an oligonucleotide (*Sal* I–*Hin*dIII–*Xba* I–N_{35}–*Eco*RI–*Bam*HI–*Xho* I) in which the *Pst* I site adjacent to the random 35-bp center was changed to *Xba* I to remove the bias introduced by the *Pst* I site (14); "com" indicates that the complementary sequence (with respect to the cloning sites) is shown.

Oligonucleotides. The multicomponent binding sites used to measure relative affinities by EMSA and relative promoter strengths in transient assays were assembled from the following oligonucleotides. Only the strand containing the consensus site as indicated in Figs. 1 and 5 is shown, and is designated U or L as convenient:

Oligonucleotide	5′ → 3′ sequence
MGNconU	GACTTGCACCTGTTGTT
MGNmutU	GACTTGTACCGGTTGTT
HeadTailU	GTTGTTCTGCACCTGTTACCCC
HeadHeadU	GTTGTTACAGGTGCAACCC
NF1conU	GTTGTTTTGGCACGGAGCCAAG
COMP1conU	TTTGATTGACAACAAC
MEF2conU	GATGCTATAAATAGACTTG

These oligonucleotides contained 6 bp of overlapping sequences, so that the multicomponent sites could be constructed by annealing and extending pairs of oligonucleotides.

Luciferase Reporter Gene Assays. pGUP.PA.8 (36) contains the human HSP70 basal promoter upstream of luciferase coding sequence (22). Consensus binding site oligonucleotides were cloned as blunt fragments into the *Sma* I site 5′ to the basal promoter of pGUP.PA.8. C2C12 myoblasts were transfected by electroporation at a density of 1.5×10^7 cells per ml of complete medium using a Bio-Rad Gene Pulser apparatus set at 270 V/960 μF. Each transfection included 10 μg of reporter plasmid, 5 μg of pCMV-lacZ, and 30 μg of sonicated salmon sperm DNA. Transfected myoblasts were plated at high density (2.5×10^5 per cm^2) in 2% defined supplemented calf serum (DSCS) to induce differentiation or at low density (3.2×10^4 per cm^2) in 20% DSCS to maintain the cells as myoblasts. Cells harvested at 24, 55, or 72 h following transfection were assayed for luciferase (22) and β-galactosidase (23). Luciferase activities were normalized to the β-galactosidase assay and are reported in comparison to the basal activity of the parental pGUP.PA.8 vector.

RESULTS

Interaction of Myogenin with NF1, MEF2, and COMP1. The sequences of individual isolates cloned following six cycles of CASTing were examined by EMSA and grouped according to the observed patterns of bands. Fig. 1 shows the consensus binding sites obtained when the sequences within three of these groups were aligned. These correspond to binding sites for a member of the NF1 family (24, 25), MEF2 (26), and another factor, which we have designated COMP1.

Direct comparisons of the affinity and transactivating ability of CASTing-derived sequences are difficult since the sequence of the binding sites varies between each different CASTing-derived clone. We thus synthesized oligonucleotides such that wild-type or mutant myogenin binding sites could be combined

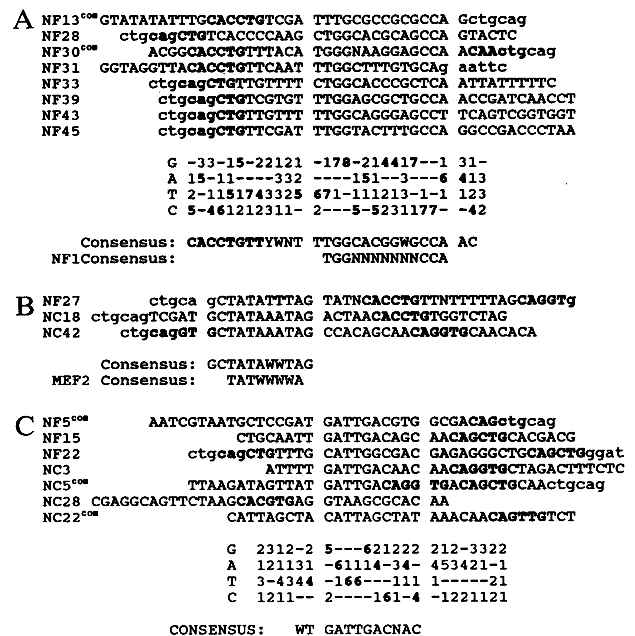

FIG. 1. Alignment of CASTing sequences exhibiting NF1, MEF2, or COMP1 patterns on EMSA. Myogenin binding sites are shown in bold letters. Nucleotides that are constrained due to their presence in restriction sites flanking the degenerate core of the original oligonucleotide are shown in lowercase letters. W = A/T; Y = G/T/C. Minimum estimates of the probability of these sites occurring by random chance are provided by the following calculations. (*A*) Sequences showing NF1 patterns on EMSA. Four sequences (NF28, NF30, NF43, and NF45) contain a 4 of 5 match to the sequence TTGGC followed by a 3 of 4 match to the sequence ACGG, then a W followed by a 4 of 5 match to GCCAA. The probability of a 4 of 5 match is 1 in $4^5 \div (4 \times 5) = 51$, that of a 3 of 4 match is 1 in $4^4 \div (4 \times 4) = 16$, and that of a W is 1 in 2. The probability of finding 4 such sequences is thus 1 in $4 \times 51 \times 16 \times 2 \times 51 = 1{:}3.3 \times 10^{-5}$. These sequences were all found within five different positions (starting between 10 and 14 bases from the center of a myogenin binding site). Of the 60 sequences analyzed, there were thus $60 \times 5 = 300$ possible positions. The probability of finding these 4 sequences within this spacing constraint is thus $300 \div 3.3 \times 10^5 = 0.9 \times 10^{-3}$. (*B*) Sequences showing MEF2 patterns on EMSA. The probability of finding three perfect matches to GCTATAWWTAG is 1 in $3 \times 4^9 \times 2^2 = 1{:}3.1 \times 10^6$. These sequences started within six different positions (starting between 18 and 23 bases from the center of a myogenin binding site). There were thus $60 \times 6 = 360$ possible positions. The probability of finding these 3 sequences within this spacing constraint is thus $360 \div 3.1 \times 10^6 = 1.2 \times 10^{-4}$. (*C*) Sequences showing COMP1 patterns on EMSA. The probability of finding 4 sequences (NF5, NF15, NC3, and NC5) with a perfect match to WTGATTGAC is 1 in $4 \times 2 \times 4^8 = 1{:}5.2 \times 10^5$. These sequences started within three positions (starting between 17 and 19 bases from the center of a myogenin binding site). There were thus $60 \times 3 = 180$ possible positions. The probability of finding these 4 sequences within this spacing constraint is thus $180 \div 5.2 \times 10^5 = 3.4 \times 10^{-4}$.

with each of the cooperating factors. Fig. 2 demonstrates that the relative affinity of myogenin is increased by the presence of either a NF1 or a COMP1 site, whereas Fig. 3 shows that myogenin interacts synergistically with both of these sites in promoting transcription from a reporter plasmid upon transfection into differentiating muscle cells.

An oligonucleotide containing the (A/T)-rich region from CASTing sequence NC18 was competitively inhibited by the MEF2 site from the creatine kinase enhancer but not by (A/T)-rich sites such as the CArG-box [CA-richG: CC(A/T)$_6$GG; (27)] (data not shown). A strong signal due to MEF2 binding to its site occurred at the same approximate mobility as the myogenin–NF1 and myogenin–COMP1 complexes. This probably prevented observation of a new band due to the

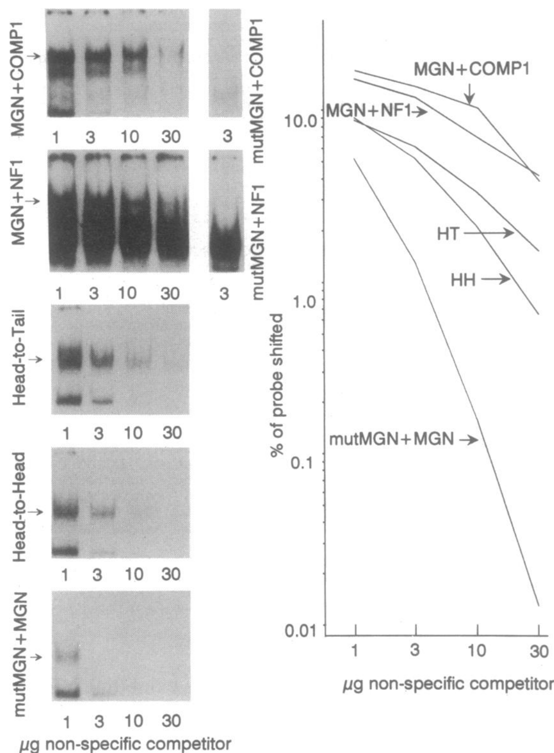

FIG. 2. Cooperative binding to multicomponent sites. The relative affinity of different complexes was examined by determining their formation in the presence of increasing amounts (1, 3, 10, and 30 μg) of nonspecific competitor (sonicated salmon sperm DNA). Synthetic probes contained the consensus binding sites for NF1, COMP1, or myogenin paired with either a myogenin or a mutant myogenin binding site. Specific myogenin-containing complexes are indicated by arrows. These complexes could be "double-shifted" by an antimyogenin antibody (data not shown), were dependent on a wild-type myogenin binding site (far right lanes), and, in the case of COMP1 and NF1, could be competitively inhibited by oligonucleotides containing COMP1 or NF1 sites but not irrelevant oligonucleotides (data not shown). The intensity of each complex was quantified as a fraction of free probe by using a Molecular Dynamics PhosphorImager and is plotted as a function of competitor DNA on the right. Multicomponent sites exhibited a higher relative affinity than a single myogenin binding site (mutMGN+MGN). The paired myogenin binding site probes contain a spacing of 14 nucleotides between the centers of their CANNTG motifs, which is optimal for a head-to-tail (HT) orientation and unfavorable for a head-to-head (HH) orientation (see Fig. 4).

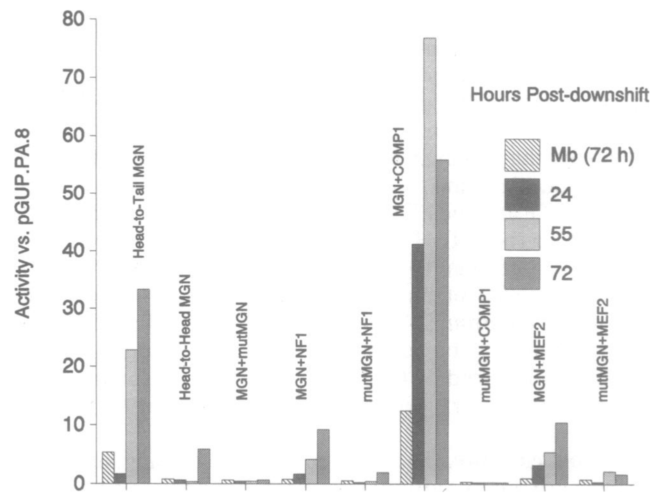

FIG. 3. Transcriptional activity of multicomponent sites. Synthetic consensus binding sites cloned into a luciferase reporter construct (pGUP.PA.8) were cotransfected into C2C12 mouse myoblasts along with a lacZ control plasmid. Activities were first normalized to β-galactosidase activity and then expressed as a fold increase over that obtained with the basal promoter alone. Since the greatest activity in myoblasts was seen at 72 h, when some small myotubes were present, the myoblast data represents an overestimate of the actual expression in "undifferentiated" cells. All of the sites interacted synergistically. The orientation of paired myogenin sites that is preferred at a spacing of 14 nucleotides (head-to-tail) gave a much greater activity than the unfavored orientation (head-to-head).

formation of a myogenin–MEF2 complex, which thus prevented our ability to directly demonstrate that MEF2 increased the affinity of myogenin binding. However, Fig. 3 shows that MEF2 and myogenin sites do interact synergistically in promoting transcription in transfection assays.

Interaction of Myogenin with Other CANNTG Binding Factors. Approximately half of the sequences contained two CANNTG motifs. Fig. 4 *Upper* shows these sequences arranged in order of increasing spacing between these sites. The myogenin consensus sequence is highly asymmetric, and most of the individual sites could be assigned an orientation. An orientation pattern emerged when the number of bases between the centers of the CANNTG motifs is plotted along the DNA double helix (Fig. 4 *Lower*). DNA undergoes a full turn approximately every 10 bp. The groupings of same versus opposite orientations alternate, such that each full turn results in the reappearance of the same group. A short transition zone, in which both groups are permitted, occurs at the borders (at spacings of 15, 16, and 20 bases). The relative positions of the two binding sites on a space-filling representation of the double helix are shown in the bottom two illustrations in Fig. 4, in which the DNA has been rotated so that the second binding sites are positioned on the surface facing out of the page. One set of interactions occurs when the spacing is one-third greater than integral (same orientation pairs at an average of 1.3 and 2.3 turns), and another set of interactions occurs at one-third less than integral (opposite orientations at an average of 1.7 and 2.7 turns). This results in the pairs of factors being able to interact (directly or indirectly) on the same side of the DNA regardless of the apparently continuous nature of the spacing.

Fig. 5 presents an alignment of the sequences based upon head-to-tail orientations (groups I and II), head-to-head (groups III and IV), and tail-to-tail (groups V and VI). The bHLH binding sites in groups I–III are very similar to each other and to the consensus site present in those CASTing sequences containing only a single myogenin binding site (TTGCACCTGTTNNTT). In contrast, the sites in groups IV–VI are quite different, suggesting that a different heterodimer may be present in each group.

Oligonucleotides containing a 14-bp spacing between paired myogenin binding sites were synthesized so that the same sites were present in what should be a favorable (head-to-tail) versus unfavorable (head-to-head) orientation. The predictions of Fig. 4 were confirmed by showing that the favored head-to-tail orientation showed an increased affinity (Fig. 2) and increased transactivation capacity (Fig. 3) compared to the head-to-head orientation at this spacing.

DISCUSSION

CASTing for multicomponent complexes indicates that myogenin can cooperatively interact, directly or indirectly, with many different factors: one member of the NF1 family, MEF2, COMP1, and perhaps as many as five other bHLH factors or complexes. These interactions not only increase the affinity of myogenin binding but also synergistically increase the transcription of reporter constructs transiently expressed in differentiating myotubes. The data also demonstrate that there are substantial spacing constraints when

```
      Spacing                                                           Orientation
NC26      7       GTCACTTTG CACCTG T CACTTG AATTTAGTATTGA                 →  •
NC55      8             ctg cagCTG TT CAGCGT GTATATCGGGGATTTATGTATAT      →  •
NF3com    9         AAGGGCG CACCTG TGA CATTTG GCACGTTCCAGA                →  •
NF11     10             ctg cagTTG GTTG CACCTG TTCGTTTTTCTGTGAGGCTG       •  →
NF38     11            GCAT CATCTG GTTTG CACCTG TTCTTGTTTGAAAT            •  →

NC54     12               C CAGCTG TTGCGC CACCTG TTGTCTAGGGATTTAA         →  →
NF25     13 AATTCTGTGTTTTG CACCTG TTCTCGG CAGTTg gatcc                    →  •
NC8      13             ctg cagCTG TTGGTTG CACCTG TCGTTTTTCTGTGAGGCTG     →  →
NF'6     14      ACCCGGCACAC CACCTG TCGGGTTG CACCTG TCAC                  →  →
NC6      14             ctg cagCTG TTAGTTTG CACCTG TTGTTTTTGGATATAGGG     →  →
NF32com  15    CTCTTGGCTATTG CACCTG TTCTTATCG CACctg cag                  →  →

NC9com   15 CTCCGGTCCCTTTG CACCTG TTGTGAAAA CAGctg cag                    →  ←
NC16     16            ACGC CACCTG TTTAATTTG CACCTG TTGTTCTTC             →  →
NF27     16 CTATATTTAGTATN CACCTG TTNTTTTAG CAGGTg gatcc                  →  ←

NF'4     17               A CAGGTG CTCCCTGCCTC CACCTG TTAGG               ←  →
NF'12    17          AGAAAA CAGGTG CTCCATACCTC CACCTG TACGTA              ←  →
NF'17    17             GAA CAGGTG CAAGAACGTAC CACCTG TTCCGGCCG           ←  →
NF1      18             ctg cagTTG TTCTGTTCGGAA CATATG CTGTGTGCGGTgga     →  ←
NF'8     20              AA CAGGTG CTCCTTCAAGGCTC CACCTG TTGGCCC          ←  →
NF'7     20              AG CAGCTG TTCTTAGACGATAA CAGGTG CCCTGCA          →  ←

NF6      20             ctg cagTTG TTTATGCNCGGAAC CAGCTG TTCCTAGA         →  →
NF16     21             ctg cagGTG TTGTATCAATTTTTG CACCTG TTACTCCAGG      →  →
NC42com  25         TGTGTTG CACCTG TTGCTGTGGCTATTTATAG CACctg cag         →  →
NF35     25             ctg cagCTG TCTTTACGTGGCTGATGCA CACCTG TTACGTA     →  →
NF30     27             ctg cagTTG TTGGCTCCTTNCCCATGTAAA CAGGTG CCGT      →  ←
NF'11    28                 CACCTG TCGTAGCTCAGGGGTGGAACAA CAGCTG T        →  ←
NF22     30             ctg cagCTG TTTGCATTGGCGACGAGAGGGCTG CAGCTG gg     →  →
```

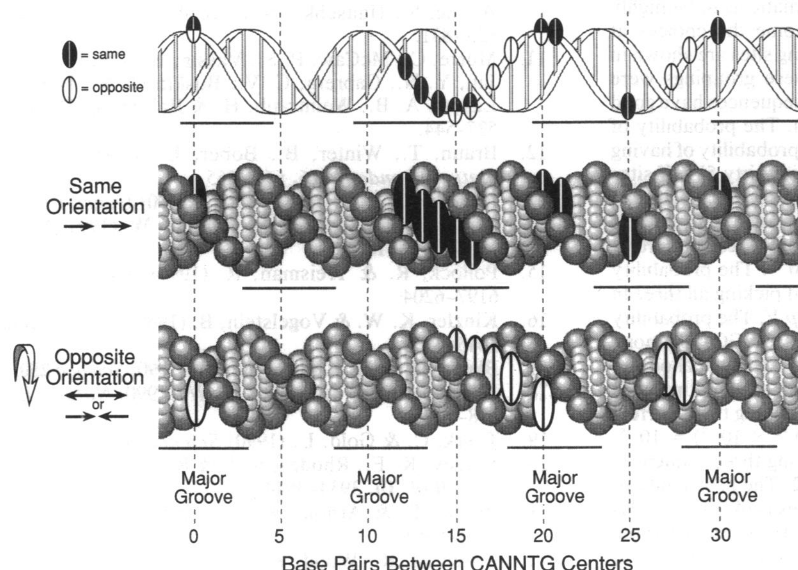

FIG. 4. Spacing between paired CANNTG motifs obtained by CASTing. (*Upper*) Sequences containing paired CANNTG motifs are arranged in order of increased spacing. The number of bases between the centers of the CANNTG motifs is shown on the left. The myogenin heterodimer consensus site (TTGCACCTGTT) has a 5′ → 3′ orientation. The orientation of each site (indicated by arrows) was determined by inspection, with a weighted importance given to first a 3′-TT, then a central CC, and finally a 5′-TTG. Sequences in which an orientation could not be determined are shown by a filled circle (•). (*Lower*) The positions of these sites along various representations of DNA are illustrated, in which the first member of each pair is located at position 0 and the second member is an appropriate number of base pairs away. The DNA molecule at the very bottom has been rotated to bring the opposite surface into view. The complete sequence is shown for each isolate other than NC9 and NF27, in which one to three bases were eliminated from one end to reduce the size of the figure.

myogenin is one of the proteins binding to paired bHLH binding sites. The effect of these constraints is to position one type of interaction (head-to-tail) an average of one-third more than an integral number of turns apart on the DNA and a second set (head-to-head and tail-to-tail) one-third less than an integral number of turns, so that each type of interaction occurs on the same face of the double helix.

Since the myogenin half-site and the MyoD half-site (17) are apparently the same, we cannot conclude that the factors binding to groups I–III of Fig. 5 are the same, in spite of the similarity of their binding sites. The present data suggest that at least four different bHLH complexes (one for groups I–III and one each for groups IV–VI) are involved in cooperative interactions involving myogenin on paired binding sites. The muscle bHLH factors can form heterodimers with various members of the E-protein gene family [E12/E47, E2-2, HEB/HTF4 (21, 28–30)], and our results suggest that several additional members of this or a related family remain to be identified in muscle extracts.

Purified MyoD has been found to bind cooperatively to the paired CANNTG motifs present in the creatine kinase enhancer (31), at spacings that largely correspond to the transition zones of Fig. 4, where either orientation is permitted. The present study establishes that myogenin can also participate in cooperative interactions involving paired CANNTG binding sites.

NF1 represents an abundant and ubiquitously expressed family of proteins, though the expression of some members is tissue-restricted (32, 33). The central core of the NF1 consensus binding site (TGG[N_{6-7}]CCA) displays considerable flexibility in supporting general NF1 family binding, whereas the motif associated with myogenin complexes contains a highly specified palindromic core (TTGGCACGGWGCCA). This sequence is similar to one of two consensus binding sites found using a similar reiterative procedure for the product of the *ski* oncogene (TGGCANNNTGCCA) (C. Richmond and E. Stavnezer, personal communication). *ski* expression induces myogenesis in quail embryo fibroblasts and the expression of MyoD and myogenin (34, 35). The similarity between the myogenin-associated NF1 site and the *ski* binding site raises the intriguing possibility that *ski* and the muscle bHLH factors might interact to cooperatively effect muscle differentiation.

The present data do not address the question of whether or not NF1, MEF2, or COMP1 could interact with myogenin if widely separated on the DNA. There may also be many interactions excluded from the present collection because they require a greater spacing. It should also be noted that since CASTing is biased toward the highest affinity sites,

```
Head to Tail:                  Group I                    Group II
NF6                      ctg cagTTG TTTATG          AAC CAGCTG TTCCTA
NF16                     ctg cagGTG TTGTAT          TTG CACCTG TTACTC
NC-42                    TTG CACCTG TTGCTG              CACctg cag
NF35                     ctg cagCTG TCTTTA          GCA CACCTG TTACGT
NF22                     ctg cagCTG TTTGCA          CTG CAGCTG ggat
Tentative Consensus:      TG CACCTG TTTNT            TG CACCTG TTACT

Head to Head:                  Group III                  Group IV
NC9                      TTG CACCTG TTGT            ctg cagCTG TTTTC
NF27                     ATN CACCTG TTNTTT        ggatc cACCTG CTAA
NF1                      ctg cagTTG TTCTGT       ACACAG CATATG TTCCGG
NF'7                      AG CAGCTG TTCTTA       GCAGGG CACCTG TTATCG
NF30                     ctg cagTTG TTGGC          ACGG CACCTG TTTACC
NF'11                      A CAGCTG TTGTTC          cag CACCTG TCGTA
Tentative Consensus:      TG CANCTG TTGTTT        CACGG CACCTG TTNTCG

Tail to Tail:                  Group V                    Group VI
NF'4                  AGGGAG CACCTG T             GCCTC CACCTG TTAGG
NF'12                 ATGGAG CACCTG TTTTCT        ACCTC CACCTG TACGTA
NF'8                  AAGGAG CACCTG TT            GGCTC CACCTG TTGGCC
NF'17                 ACGTAC CACCTG TTCCGG        TCTTG CACCTG TTC
Tentative Consensus:ANGGAG CACCTG TT              GCCTC CACCTG TTCG

Overall Groupings:
                    I:  TG CACCTG TTTNT    IV: CACGG CACCTG TTNTCG
                   II:  TG CACCTG TTACT     V:ANGGAG CACCTG TT
                  III:  TG CANCTG TTGTTT   VI: GCCTC CACCTG TTCG
```

FIG. 5. Consensus sequence of paired CANNTG sites. Paired sites from Fig. 4 are grouped according to orientation. The DNA strand containing the CTGTT muscle bHLH half-site at the 3′ end is shown. The consensus sequences are listed as tentative since relatively few examples of each group are available. The differences between the four types of 5′ half-sites are so dramatic as to be highly significant in spite of the tentative nature of some the choices at specific positions. Bases contributed by the cloning sites are shown in lowercase letters. Minimum probabilities for these groupings were calculated as follows. *Group IV*. Three of the six sequences have good matches to CACGG (the rest are uninformative). The probability of having a 4 of 5 match is $1{:}4^5 \div (4 \times 5) = 1{:}51$. The probability of having three sequences with a 4 of 5 match to CACGG in thirty 5′ half-sites is $30 \div (3 \times 51) = 0.2$. The number of ways three yes/no choices can be made in six tries is 20. The probability of choosing the three best matches to CACGG out of thirty 5′ half-sites in six tries is approximately $(3/30) \times (2/29) \times (1/28) \times 20 = 4.9 \times 10^{-3}$. The probability of finding three sequences with a 4 of 5 match and picking all three in six tries is thus $(4.9 \times 10^{-3}) \times (0.2) = 10^{-3}$. *Group V*. The probability of finding three sequences with a perfect match to ANGGAG among thirty 5′ half-sites is $30 \div (4^5 \times 3) = 10^{-2}$. The probability of choosing these three sequences in four tries from thirty 5′ half-sites is 1 in $(3/30) \times (2/29) \times (1/28) \times 4 = 10^{-3}$. The probability of finding three perfect matches and picking all three in four tries is $(10^{-2} \times 10^{-3}) = 10^{-5}$. *Group VI*. As in group IV, the probability of having three sequences with a 4 of 5 match out of thirty 5′ half-sites is 0.2. The probability of picking the four best matches out of 30 in four tries is $(4/30) \times (3/29) \times (2/28) \times (1/27) = 3.6 \times 10^{-5}$. The probability of finding three sequences with a 4 of 5 match and picking the best four out of 30 in four tries is $0.2 \times (3.6 \times 10^{-5}) = 0.7 \times 10^{-5}$.

there may be many other multicomponent complexes involving myogenin that were not retrieved in these experiments because of their lower affinity. There is no *a priori* reason to assume that the highest affinity interactions have more biological importance than lower affinity ones when dealing with the complicated combinatorial processes involved in muscle determination, differentiation, and the modulation of expression in mature muscle fibers.

Tissue-specific enhancers are complex regulatory elements that contain multiple binding sites for cooperating and competing regulatory factors. Previous studies using CASTing approaches to define DNA binding sites have largely used purified proteins and thus were limited in the information they could obtain. The present use of crude nuclear extracts as a source of multiple components, a degenerate core large enough to accommodate several different binding sites, and a monoclonal antibody to provide specificity for particular multicomponent complexes represents a major advance that permits the identification of multicomponent complexes involved in tissue-specific transcription and modulation of gene activity.

Oligonucleotides to previously identified sites were kindly provided by Jean Buskin (the MEF1 oligonucleotide to the 3′ CANNTG site of the creatine kinase enhancer), Eric Olson (the MEF2 site from the creatine kinase enhancer), and Margaret Buckingham (the CArG box from the cardiac α-actin promoter). This work was supported by the National Institutes of Health (Grant AGO1228) (W.E.W.), the Muscular Dystrophy Association of America (W.E.W.), and a postdoctoral fellowship from the Medical Research Council of Canada (W.D.F.).

1. Wright, W. E., Sassoon, D. A. & Lin, V. K. (1989) *Cell* **56**, 607–617.
2. Edmondson, D. G. & Olson, E. N. (1989) *Genes Dev.* **3**, 628–640.
3. Davis, R. L., Weintraub, H. & Lassar, A. B. (1987) *Cell* **37**, 879–887.
4. Braun, T., Buschhausen-Denker, G., Bober, E., Tannich, E. & Arnold, H. H. (1989) *EMBO J.* **8**, 701–709.
5. Rhodes, S. J. & Konieczny, S. F. (1989) *Genes Dev.* **3**, 2050–2061.
6. Miner, J. H. & Wold, B. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 1089–1093.
7. Braun, T., Bober, E., Winter, B., Rosenthal, N. & Arnold, H. H. (1990) *EMBO J.* **9**, 821–831.
8. Murre, C., McCaw, P. S. & Baltimore, D. (1989) *Cell* **56**, 777–783.
9. Farmer, K., Catala, F. & Wright, W. E. (1992) *J. Biol. Chem.* **267**, 5631–5636.
10. Lassar, A. B., Buskin, J. N., Lockson, D., Davis, R. L., Apone, S., Hauschka, S. D. & Weintraub, H. (1989) *Cell* **58**, 823–831.
11. Murre, C., McCaw, P. S., Vassin, H., Caudy, M., Jan, L. Y., Jan, Y. N., Cabrera, C. V., Buskin, J. N., Hauschka, S. D., Lassar, A. B., Weintraub, H. & Baltimore, D. (1989) *Cell* **58**, 537–544.
12. Braun, T., Winter, B., Bober, E. & Arnold, H. H. (1990) *Nature (London)* **346**, 663–665.
13. Brennan, T. J. & Olson, E. N. (1990) *Genes Dev.* **4**, 582–595.
14. Wright, W. E., Bender, M. & Funk, W. (1991) *Mol. Cell. Biol.* **11**, 4104–4110.
15. Pollock, R. & Treisman, R. (1990) *Nucleic Acids Res.* **18**, 6197–6204.
16. Kinzler, K. W. & Vogelstein, B. (1989) *Nucleic Acids Res.* **17**, 3645–3653.
17. Blackwell, T. K. & Weintraub, H. (1990) *Science* **250**, 1104–1110.
18. Ellington, A. D. & Szostak, J. W. (1990) *Nature (London)* **346**, 818–822.
19. Tuerk, C. & Gold, L. (1990) *Science* **249**, 505–510.
20. Yutzey, K. E., Rhodes, S. J. & Konieczny, S. F. (1990) *Mol. Cell. Biol.* **10**, 3934–3944.
21. Braun, T. & Arnold, H. H. (1991) *Nucleic Acids Res.* **19**, 5645–5651.
22. DeWet, J. R., Wood, K. V., DeLuca, M., Helinski, D. R. & Subramani, S. (1987) *Mol. Cell. Biol.* **7**, 725–737.
23. Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab., Cold Spring Harbor, NY), 2nd Ed.
24. Gronostajski, R. M. (1986) *Nucleic Acids Res.* **14**, 9117–9131.
25. Nowock, J., Borgmeyer, U., Puschel, A. W., Rupp, R. A. W. & Sippel, A. E. (1985) *Nucleic Acids Res.* **13**, 2045–2061.
26. Gosset, L. A., Kelvin, D. J., Sternberg, E. A. & Olson, E. M. (1989) *Mol. Cell. Biol.* **9**, 5022–5033.
27. Minty, A. & Kedes, L. (1986) *Mol. Cell. Biol.* **6**, 2125–2136.
28. Lassar, A. B., Davis, R. L., Wright, W. E., Voronova, A., Baltimore, D. & Weintraub, H. (1991) *Cell* **66**, 305–315.
29. Hy, J.-S., Olson, E. N. & Kingston, R. E. (1992) *Mol. Cell. Biol.* **12**, 1031–1042.
30. Zhang, Y., Babin, J., Feldhaus, A. L., Singh, H., Sharp, P. A. & Bina, M. (1991) *Nucleic Acids Res.* **19**, 4555.
31. Weintraub, H., Davis, R., Lockshon, D. & Lassar, A. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 5623–5627.
32. Lichsteiner, S., Wuarin, J. & Schibler, U. (1987) *Cell* **51**, 963–973.
33. Paonessa, G., Gounari, F., Frank, R. & Cortese, R. (1988) *EMBO J.* **7**, 3115–3123.
34. Colmenares C. & Stavnezer, E. (1989) *Cell* **59**, 293–303.
35. Colmenares C. & Stavnezer, E. (1991) *Mol. Cell. Biol.* **11**, 1167–1170.
36. Bassel-Duby, R., Hernandez, M. D., Gonzalez, M. A., Krueger, J. K. & Williams, R. S. (1992) *Mol. Cell. Biol.*, in press.