Dual DNA Binding Specificity of ADD1/SREBP1 Controlled by a Single Amino Acid in the Basic Helix-Loop-Helix Domain

JAE B. KIM,[1] GERALD D. SPOTTS,[1] YUAN-DI HALVORSEN,[1] HSIU-MING SHIH,[2]
TOM ELLENBERGER,[3] HOWARD C. TOWLE,[2] AND BRUCE M. SPIEGELMAN[1]*

*Dana-Farber Cancer Institute and Department of Cell Biology, Harvard Medical School,[1] and Department
of Biological Chemistry and Molecular Pharmacology, Harvard Medical School,[3] Boston,
Massachusetts 02115, and Department of Biochemistry and Institute of Human
Genetics, University of Minnesota, Minneapolis, Minnesota 55455[2]*

Adipocyte determination- and differentiation-dependent factor 1 (ADD1), a member of the basic helix-loop-helix (bHLH) family of transcription factors, has been associated with both adipocyte differentiation and cholesterol homeostasis (in which case it has been termed SREBP1). Using PCR-amplified binding analysis, we demonstrate that ADD1/SREBP1 has dual DNA sequence specificity, binding to both an E-box motif (ATCACGTGA) and a non-E-box sequence previously shown to be important in cholesterol metabolism, sterol regulatory element 1 (SRE-1; ATCACCCCAC). The ADD1/SREBP1 consensus E-box site is similar to a regulatory sequence designated the carbohydrate response element, defined by its ability to regulate transcription in response to carbohydrate in genes involved in fatty acid and triglyceride metabolism in liver and fat. When expressed in fibroblasts, ADD1/SREBP1 activates transcription through both the carbohydrate response E-box element and SRE-1. Substitution of an atypical tyrosine in the basic region of ADD1/SREBP1 to an arginine found in most bHLH protein causes a restriction to only E-box binding. Conversely, substitution of a tyrosine for the equivalent arginine in another bHLH protein, upstream stimulatory factor, allows this factor to acquire a dual binding specificity similar to that of ADD1/SREBP1. Promoter activation by ADD1/SREBP1 through the carbohydrate response element E box is not sensitive to the tyrosine-to-arginine mutation, while activation through SRE-1 is completely suppressed. These data illustrate that ADD1/SREBP1 has dual DNA sequence specificity controlled by a single amino acid residue; this dual specificity may provide a novel mechanism to coordinate different pathways of lipid metabolism.

The basic helix-loop-helix (bHLH) transcription factors regulate gene expression by binding to specific DNA sequences. The basic domain of these proteins controls DNA binding to sites with the consensus sequence CANNTG. This consensus sequence is referred to as the E-box motif and is present in the regulatory regions of many tissue-specific genes (6, 8, 13, 27, 29). The function of the basic domain in DNA binding has been illustrated by mutations that disrupt the interaction with DNA but not oligomerization (13, 34). The Id protein, which has no basic domain, does not bind DNA but can act as a dominant suppressor of DNA binding of certain other bHLH proteins (3, 16). The HLH domain mediates homo- and hetero-oligomerization through two amphipathic α helices connected by a variable loop region (1, 3, 13, 29). The various bHLH proteins can be divided into at least three groups (30). These include the broadly expressed class A proteins (E12, E47, E2-2, and daughterless) (11, 22), the tissue-specific class B proteins (MyoD, myogenin, MRF4, and achaete-scute) (9, 14, 36), and class C proteins, which feature a tandem arrangement of bHLH and leucine zipper (LZ) motifs (c-Myc, Max, upstream stimulatory factor [USF], AP4, TFE3, and TFEB) (2, 5, 7, 10, 21, 24).

Adipocyte determination- and differentiation-dependent factor 1 (ADD1) is a member of the bHLH-LZ family of transcriptional activators (33). We cloned ADD1 by screening an adipocyte cDNA library with an oligonucleotide probe con-

taining an E-box motif (33). ADD1 mRNA is elevated in determined preadipocytes and is increased further during the process of adipose differentiation. Independently, Brown and colleagues purified and cloned the human homolog of ADD1, which they termed sterol regulatory element-binding protein 1 (SREBP1), which binds the non-E-box sequence sterol regulatory element 1 (SRE-1; ATCACCCCAC), a key regulatory element in the promoter of several genes involved in cholesterol homeostasis (35, 37). The SRE-1 sequence enhances transcription in sterol-depleted cells, and its activity is abolished when sterols accumulated (20). SREBP1 regulates the transcription of the low-density lipoprotein (LDL) receptor gene and 3-hydroxy-3-methylglutaryl coenzyme synthase gene through the SRE-1 motif (20, 37). However, these authors note that although SREBP1 is a bHLH factor, it does not bind to E boxes (37).

To explore the issue of ADD1/SREBP1 DNA binding specificity, we have selected the optimal DNA binding sites of ADD1/SREBP1 from a pool of random oligonucleotides. We demonstrate that the ADD1/SREBP1 protein can indeed bind both E-box (ATCACGTGT) and non-E-box (ATCACCCCAC; SRE-1) motifs, and such a dual DNA binding specificity of ADD1/SREBP1 is the result of an atypical tyrosine residue in the conserved basic domain. In addition, we demonstrate here that ADD1/SREBP1 can activate transcription through the carbohydrate response element E box found in the fat- and liver-specific S_{14} gene. These results suggest that the ADD1/SREBP1 homodimer has dual DNA binding specificity and is likely to regulate the different sets of genes involved in lipid metabolism.

* Corresponding author. Mailing address: Dana-Farber Cancer Institute, 44 Binney St., Boston, MA 02115. Phone: (617) 632-3567. Fax: (617) 632-4655. Electronic mail address: Bruce_Spiegelman@dfci.harvard.edu.

MATERIALS AND METHODS

Optimal binding sites amplification. To identify the DNA sequences that were tightly bound by ADD1/SREBP1, bacterial fusion protein was incubated at a limiting concentration in a standard binding reaction mixture (23, 26) containing ^{32}P-labeled random oligonucleotides, flanked on either side by a known sequence to facilitate PCR amplification and subsequent cloning of the recovered DNA. The DNA sequence of the random oligonucleotide was 5′-AGTGGAAT TCGCGAAGATGGGCGTGN$_{16}$GGAGACCAGTAGGCATGCGACCC-3′. The oligonucleotide was labeled with [^{32}P]dCTP by annealing a complementary oligonucleotide (5′-GGGTCGCATGCCTACTGGTCTCC-3′) and extension of the annealed primer as described previously (6). To isolate the bound DNA, protein-DNA complexes were separated by electrophoresis, excised from the gel, and eluted from the acrylamide as described previously (6, 15). Eluted DNA was amplified by PCR with primers that were complementary to the nonrandom flanking sequence. PCR was performed under standard conditions (28). Five additional cycles of DNA-protein binding, elution, and PCR amplification were performed. After the sixth round of amplification, the PCR products were cloned into plasmid pCRII (Invitrogen) and subjected to dideoxy sequencing (33).

Fusion protein production and EMSA. Plasmid pTrcHis C (Invitrogen) was used for expression of ADD1/SREBP1 protein in *Escherichia coli*. cDNA encoding amino acid residues including the bHLH-LZ domain (residues 284 to 403 [ADD1-S] and residues 6 to 403 [ADD1-L]) were cloned by PCR with *Bam*HI and *Eco*RI ends and fused in frame with a six-histidine moiety present in plasmid pTrcHis C. USF (21) was overexpressed in *E. coli* as described above. USF cDNA fragments were also engineered at both ends to clone the *Bam*HI and *Eco*RI sites of plasmid pTrcHis C with an in-frame fusion. USF-S and USF-L contained amino acid residues 175 to 310 and 2 to 310, respectively. The fusion protein expression and affinity purification were performed under conditions recommended by the manufacturer (Qiagen). Briefly, following induction of mid-logarithmic-phase cultures of *E. coli* with 1 mM isopropyl-β-D-thiogalacto-pyranoside (IPTG), the cells were lysed with 6 M guanidine hydrochloride (pH 8.0), and fusion proteins were purified by passing the cleared lysate over a nickel chelate affinity resin. Fusion proteins were eluted with 500 mM imidazole and used in electrophoretic mobility shift assays (EMSAs). For binding assays with in vitro-translated ADD1/SREBP1 protein, C-terminal truncated ADD1 cDNA containing amino acid residues 1 to 546 was cloned into plasmid pSVSPORT1 (Gibco/BRL) and translated in the TNT SP6 coupled reticulocyte lysate system (Promega). Binding reactions were performed as described previously (2, 27) with 10 mM Tris (pH 7.5), 100 mM KCl, 10 mM MgCl$_2$, 1 mM EDTA, 1 mg of poly(dI-dC), 1 mM dithiothreitol, and 10 mg of bovine serum albumin per ml. For binding assays, reactions were performed in a 20-μl volume containing purified bacterial protein (10 ng) or in vitro translation lysate (2 μl) and ^{32}P-labeled probe (0.1 pmol) with a 10-min incubation at room temperature. The EMSA was performed on a 4% polyacrylamide gel with 0.25× Tris-borate-EDTA (TBE) buffer as described previously (6, 26) and processed for autoradiography. For binding competition assays, unlabeled oligonucleotides were added prior to the addition of labeled probe. The DNA sequence of double-stranded oligonucleotides used were as follows (only one strand is shown): ABS (ADD1/SREBP1 binding site), 5′-GATCCTGATCACGTGATCGAGGAG-3′; SRE-1, 5′-GATCCTGATCACCCCACTGAGGAG-3′; and USE (upstream stimulatory elements), 5′-GATCCTGGTCACGTGGCCGAGGAG-3′. Mutant proteins that encoded the desired amino acid sequence were produced by the PCR and cloned into plasmid pTrcHis C. ADD1 (Y-320→R) mutant and USF (R-211→Y) mutant cDNAs were constructed by use of primers 5′-ATTGAG AAGCGC*CGA*CGTTCCTCTATC-3′ and 5′-GTGGAGCGTCGC*TAT*CGAG ACAAGATC-3′, respectively (only the sense strand is shown). Mutations were introduced individually at the underlined positions noted above in the primers: ADD1 (Y-320→R) CGA and USF (R-211→Y) TAT.

Transfection and CAT assay. NIH 3T3 cells were cultured in Dulbecco modified Eagle medium (Sigma) containing 10% bovine calf serum (HyClone) and transfected 1 day postconfluence by the calcium phosphate method previously described (33). Three tandem repeats of the ABS or SRE-1 oligonucleotide were cloned into pSVKS1, a chloramphenicol acetyltransferase (CAT) reporter plasmid driven by the enhancerless simian virus 40 promoter (33). cDNA of wild-type ADD1 or Y-320→R mutated ADD1 was cloned into pSVSPORT1 expression vector. The ADD1/SREBP1 (Y-320→R) mutation in an expression vector was produced by PCR as described above. The S$_{14}$ CAT construct contained four copies of carbohydrate response element (−1457 to −1428 of the rat S$_{14}$ gene) (25, 31, 32). NIH 3T3 cells were cotransfected with ABS CAT, SRE CAT, or S$_{14}$ CAT plasmid (2 μg) and ADD1/SREBP1 or ADD1/SREBP1 (Y-320→R) expression vector (2 μg) as described earlier (33). The level of CAT gene expression from each transfection was determined by measuring CAT enzyme activity. All transfection experiments were performed in duplicate, and at least three different CsCl preparations of each plasmid were tested. For the normalization of CAT assays, relative transfection efficiencies were determined by cotransfection of β-galactosidase expression plasmid pCH110 (Pharmacia LKB) and subsequent β-galactosidase assays.

RESULTS

Optimal DNA binding sites of ADD1/SREBP1. To clarify the binding specificity of ADD1/SREBP1 and to identify its optimal DNA binding sequence, the technique of selected and amplified binding sequence (SAAB) imprinting was used with a pool of completely random oligonucleotides (6). The oligonucleotide pool used for binding and PCR amplification contained 16 contiguous random nucleotides flanked on either side by a known DNA sequence to facilitate PCR amplification and subsequent cloning (see Materials and Methods). The radiolabeled random oligonucleotide pool was incubated with purified 18-kDa six-histidine–ADD1/SREBP1 fusion protein of bacterial origin, ADD1-S (see Materials and Methods). The ADD1/SREBP1 sequences of the fusion proteins include the bHLH domain and the LZ and therefore should contain the minimal structural requirements for DNA binding. After six successive rounds of selection and PCR amplifications, the final PCR products were cloned and sequenced.

As shown in Fig. 1A, 30 clones contained the E-box sequence of CANNTG, with most clones containing a CACGTG core. This CACGTG motif is a common E-box core sequence which is found in a number of promoters, including the USE within the adenovirus major late promoter (19). Other proteins with similar bHLH-LZ structure, such as Myc, Max, USF, TFE3, TFEB, and CBF1, can also bind to this CACGTG E-box sequence (4, 6, 13). The flanking sequences as well as internal core sequence of E boxes are crucial for the specific binding of each bHLH protein to its cognate E-box site (6). Like other E-box-binding bHLH proteins, the ADD1/SREBP1 protein can also bind weakly to several different E-box sites, i.e., CAC CTG, CATGTG, and CACATG sites. An alignment of the selected E-box motifs suggests sequence preferences at positions flanking these hexamers (Fig. 1A). The preferential ABS can be represented by the palindromic consensus sequence ATCACGTGAPy. An A preference at position −5 is mirrored by a pyrimidine at +5, a T preference at −4 corresponds to an A at +4, and the −1 and +1 preferences are C and G, respectively (Fig. 1A). The ADD1/SREBP1 fusion protein also selected seven clones with the non-E-box (ATCACCCCAC) motif, which has been identified as SRE-1 (Fig. 1B) (20, 35). The SRE-1 motif shows homology with a half-site of ABS (ATC AC) in the 5′ half but two direct repeats of the CAC sequence.

To ascertain whether the optimal E-box sequence (ABS) represents a true high-affinity binding site for ADD1/SREBP1 protein and to evaluate the DNA binding specificity of the ADD1/SREBP1 protein, EMSAs were performed. Bacterially produced ADD1/SREBP1 fusion protein and in vitro-translated ADD1/SREBP1 protein were analyzed in parallel experiments (Fig. 2A). The ^{32}P-labeled ABS probe was incubated with increasing levels of unlabeled competitor oligonucleotides containing the ABS, SRE-1, or USE sequence. USE, the binding site for USF, shares the E-box core sequence (CACGTG) with the ABS but has different flanking sequences (4). As illustrated in Fig. 2B, binding of ADD1/SREBP1 protein to the ABS probe was decreased by competition with excess unlabeled ABS, SRE-1, or USE oligonucleotide. However, ABS competed two- to threefold better than the SRE-1 sequence, while SRE-1 competed two- to threefold better than the USE sequence (Fig. 2B). The same results were obtained when either the SRE-1 or USE oligonucleotide was used as the probe (data not shown). These observations indicate that ADD1/SREBP1 can bind to both E-box and non-E-box (SRE-1) sequences with high affinity. Different target sequences apparently can compete for ADD1/SREBP1 protein

A E-Box (CANNTG)

```
A1          gcgtg CTGGTGT CACGTG TTT ggaga
A2           gcgtg TAT CACGTG AGCGTTC ggaga
A3          gcgtg GGTCAT CACGTG AGCA ggaga
A4         tctcc ATATAGTAT CACGTG A cacgc
A5          gcgtg TGAT CATGTG ATGGCC ggaga
A6           tctcc ACAT CATGTG ATTATG cacgc
A7          gcgtg GTAT CACGTG ACTAGC ggaga
A8         tctcc CAACGGAT CACCTG AT cacgc
A9           tctcc TGAT CACATG CCAGTA cacgc
A10        gcgtg TTGATTCAT CACGTG C ggaga
A11        tctcc TGGCACAAT CACGTG A cacgc
A12         tctcc TTATAGT CACGTG ACT cacgc
A13        tctcc ATCTAGT CACGTG ACA cacgc
A14          tctcc GCACT CACCTG ATTTG cacgc
A15        tctcc ACAGACGGT CACGTG A cacgc
A16         tctcc AACACAC CACGTG ACC cacgc
A17         gcgtg GCGAT CACGTG TATTA ggaga
A18        tctcc ACAGATGAT CACGTG A cacgc
A19         gcgtg TTCAT CACCTG AGTCA ggaga
A20        gcgtg GCGAT CAGGTG ATGCA ggaga
A21         gcgtg GGT CACGTG GTGTGCA ggaga
A22        tctcc CAAGCGGAT CACGTG A cacgc
A23        tctcc TGCACACTA CACGTG A cacgc
A24          tctcc TGAT CACATG CCATGT cacgc
A26         gcgtg GGTCAT CACGTG AGCC ggaga
A27        gcgtg CAACGGAT CACCTG AT ggaga
A28        tctcc TCTGTTCAT CACGTG A cacgc
A29        tctcc TCATGCAT CACGTG AT cacgc
A30        gcgtg TGCACAAT CACGTG AT ggaga
```

Position	-6	-5	-4	-3	-2	-1	1	2	3	4	5
G	11	5				1	22		30	1	5
A	4	23	1		30		2			21	2
T	5	1	28			2	.	30		2	11
C	10	1	1	30	27	6				3	12
	G/C	A	T	C	A	C	G	T	G	A	Py

ADD1/SREBP1 E-Box Consensus Binding Sequence:

5'-ATCACGTGA-3'

B Non-E-Box

```
S1          tctcc GATCTTGAG ATCACCCcac gc
S2          tctcc AGAGCTAGT GTCACCCcac gc
S3          tctcc TTGTGTCTG ATCACCCcac gc
S4              tctcc ATCACCCCAC CTGTCGcacgc
S5          gcgtg TCAGAG ATCACCCCAC ggaga
S6          gcgtg CTAGCTCGT GTCACCCcac gc
S7          tctcc ACGTATTGT ATCACCCcac gc
```

Position	-1	1	2	3	4	5	6	7	8	9	10
G	3	2									
A		5				7				7	
T	3		7								
C	1			7		7	7	7	7		7
	N	A	T	C	A	C	C	C	C	A	C

ADD1/SREBP1 Non-E-Box Consensus Binding Sequence:

SAAB Consensus sequence: 5'-ATCACCCCAC-3'
Published SRE-1 sequence: 5'-ATCACCCCAC-3'

FIG. 1. Nucleotide sequences of selected binding sites of ADD1/SREBP1 protein and the predicted consensus alignments. DNA sequences that bound to ADD1/SREBP1 fusion protein were selected and analyzed by DNA sequencing. The cloned random sequences are indicated in uppercase letters, and portions of the nonrandom flanking sequences are indicated in lowercase letters. (A) The 30 E-box (CANNTG) sequences obtained by SAAB analysis were selected, and potential E-box sequences are highlighted. The E-box sequences are aligned from positions −6 to +5. The E-box consensus sequence was determined by choosing the most frequent nucleotide at each position. (B) The seven non-E-box sequences obtained by simultaneous SAAB analysis were selected, and conserved non-E-box (SRE-1) sequences are highlighted. The conserved non-E-box sequences (SRE-1) were aligned from positions −1 to +10, and the non-E-box consensus sequence was determined accordingly.

in a single conformation, or if two different conformations of ADD1/SREBP1 exist, they must be in a rapid equilibrium.

Dimerization of ADD1/SREBP1. DNA-binding proteins that recognize palindromic DNA sequences commonly bind DNA as dimers. Indeed, most bHLH-LZ proteins form homodimers or heterodimers through their HLH-LZ domains in order to bind the palindromic E-box motif (3, 6, 16, 34). To determine whether ADD1/SREBP1 protein binds the E-box (ABS) motif and the non-E-box (SRE-1) motif with the same oligomerization state, two different-size ADD1/SREBP1 fusion proteins, ADD1-S (18 kDa) and ADD1-L (45 kDa), were produced in bacteria and used for EMSA (Fig. 2A). The two proteins have similar DNA binding affinities and contain all residues required for sequence-specific DNA binding. Each of these proteins was mixed with radiolabeled DNA probes, and the DNA-protein complexes were resolved by EMSA. The presence of a single intermediate is most consistent with a dimeric structure for the protein complex (Fig. 2C, lanes 3 and 6), which can bind both the E-box (ABS) sequence and the non-E-box (SRE-1) sequence (Fig. 2C). These observations imply that ADD1/SREBP1 proteins form homodimers in order to bind a DNA target site containing either an E-box (ABS) or a non-E-box (SRE-1) motif.

Dual DNA binding specificity of ADD1 is regulated by a single tyrosine residue. To elucidate the structural features that allow for the dual binding specificity of ADD1/SREBP1 protein, we compared the amino acid sequences of ADD1/SREBP1 and other bHLH proteins. In the HLH domain, similarities between family members are restricted to the conserved residues (Fig. 3) that are believed to confer the overall structure of the DNA binding domain. Mutational analysis of MyoD has shown that the basic region is required for DNA binding and muscle-specific gene activation but not for protein dimerization, which requires only the HLH motif (13). We therefore focused on the cluster of basic residues adjacent to the first amphipathic helix. Alignment of the basic domains of the bHLH proteins demonstrated a clear pattern of conserved and nonconserved residues. The bHLH proteins that have been shown to bind the CANNTG consensus site contained a conserved E-R-X-R or E-K-X-R sequence (where X represents any amino acid) in the basic region at exactly the same position (Fig. 3). However, the ADD1/SREBP1 protein has an atypical tyrosine residue at position 320 (E-K-R-Y) in the basic domain instead of the conserved arginine observed in all other known bHLH proteins. The rest of basic and HLH domains of the ADD1/SREBP1 protein are well conserved.

The role of this tyrosine residue in the DNA binding specificity of ADD1/SREBP1 was examined by replacing it with an arginine residue. The resulting ADD1/SREBP1 mutant (Y-320→R) homodimer bound to an E-box site (ABS) (Fig. 4A, lane 2) but not to the SRE-1 site (Fig. 4A, lane 5). However, when the mutant ADD1/SREBP1 (Y-320→R) protein formed dimers with the wild-type fusion protein, the resulting heterodimer was then able to bind the SRE-1 motif as well as the E-box motif (ABS) (Fig. 4A, lanes 3 and 6). Furthermore, when the tyrosine residue of ADD1/SREBP1 was substituted with alanine or phenylalanine, these mutant proteins lost their DNA binding activities (data not shown). These results suggest that the tyrosine residue in the basic domain of ADD1/SREBP1 is responsible for the dual DNA binding specificity and that one subunit of the ADD1/SREBP1 protein complex must be wild type in order to bind the SRE-1 sequence.

To determine whether the role of this unusual tyrosine residue could be generalized to other bHLH proteins, we substituted the equivalent arginine 211 with a tyrosine in USF, a member of bHLH-LZ family, which is known to recognize the
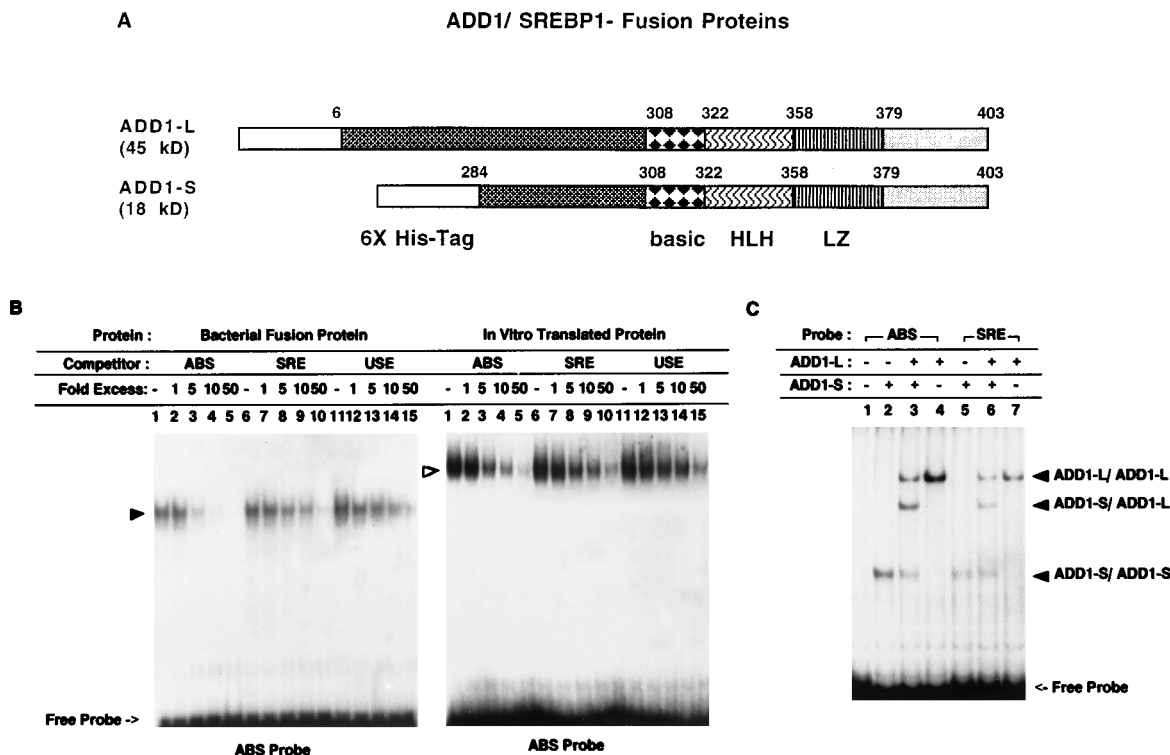
A **ADD1/ SREBP1- Fusion Proteins**

FIG. 2. Gel mobility assays of ADD1/SREBP1 protein. (A) Diagram of ADD1/SREBP1 fusion protein. The six-histidine moiety, the bHLH region, the LZ, and flanking ADD1/SREBP1 sequences are indicated. The amino-terminal amino acid of each ADD1/SREBP1 domain is numbered according to its position in full-length rat ADD1 (33). (B) Analysis of relative affinities of the ABS versus SRE-1 and USE for ADD1/SREBP1 protein. The same amount of either purified bacterial fusion ADD1/SREBP1 protein or in vitro-translated ADD1/SREBP1 protein was added to increasing amounts of unlabeled ABS oligonucleotide (lane 1 to 5), SRE-1 oligonucleotide (lane 6 to 10), or USE oligonucleotide (lane 11 to 15) as the competitor DNA. In all lanes, ^{32}P-radiolabeled ABS oligonucleotide was used as the probe. The ABS-ADD1/SREBP1 bacterial fusion protein complexes are marked by a closed arrowhead, and ABS–in vitro-translated ADD1/SREBP1 protein complexes are marked by an open arrowhead. (C) ADD1/SREBP1 bind to DNA as a dimer. Two different sizes of ADD1/SREBP1 fusion proteins were used for EMSA. ^{32}P-labeled DNA probes were prepared from an oligonucleotide containing the E-box consensus (ABS) sequence or the non-E-box consensus (SRE-1) sequence. The probes were incubated with the indicated combinations of bacterially produced short and long forms of ADD1/SREBP1 protein. The predicted protein composition of each complex is indicated.

USE motif containing CACGTG E-box sequence. The USF (R-211→Y) mutant and wild-type fusion proteins were bacterially produced, and their DNA binding specificities were assayed with the oligonucleotides containing either USE or SRE-1. As expected, wild-type USF fusion protein bound to a cognate USE site with high affinity but did not bind to the SRE-1 site (Fig. 4B). However, the equivalent USF (R-211→

Y) mutant fusion protein bound specifically to both the E-box (USE) and SRE-1 sites, similar to the ADD1/SREBP1 protein (Fig. 4B). These results confirm the crucial role of the tyrosine in dual specificity of binding and suggest that its function can be generalized to other bHLH-LZ family members.

Transcriptional activation through dual binding sites. As a first approach to determining the role that ADD1/SREBP1
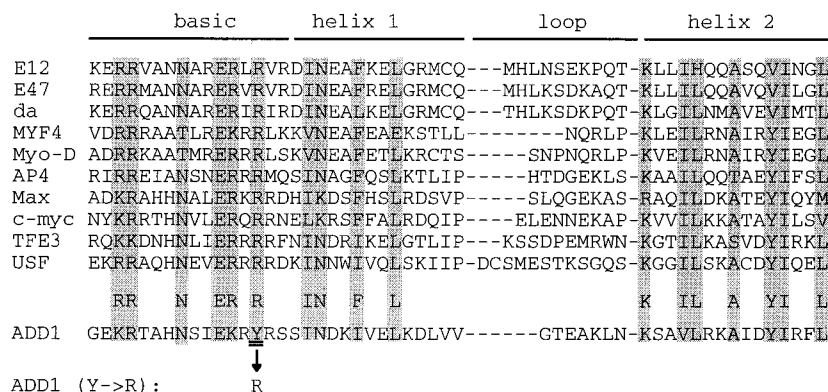

FIG. 3. Amino acid sequence alignment of bHLH proteins. The basic helix 1-loop-helix 2 domains of bHLH proteins E12, E47, daughterless and bHLH-LZ proteins Myf4, AP4, Max, c-Myc, TFE3, and USF were compared with the ADD1/SREBP1 protein. Highly conserved residues are highlighted. The atypical tyrosine residue of ADD1/SREBP1 protein is underlined.

A

Probe: ABS SRE
ADD1-L: -   -   +   +   -   +   +
ADD1(Y->R)-S: -   +   +   -   +   +   -
                1   2   3   4   5   6   7

◄ ADD1-L/ ADD1-L
◄ ADD1(Y->R)-S/ ADD1-L
◄ ADD1(Y->R)-S/ ADD1(Y->R)-S
<- Free Probe

B

Probe: U   U   S   U   S   U   U   U   U   U   U   U   S   S   S
USF-S: -   +   +   -   -   +   +   +   -   -   -   -   -   -   -
USF-L: -   -   -   -   -   -   +   +   -   +   +   -   +   +   -
USF (R-> Y)-S: -   -   -   +   +   -   -   -   +   +   -   +   +   -
         1   2   3   4   5   6   7   8   9   10 11 12 13 14

◄ USF-L/ USF-L
◄ USF(R->Y)-S/ USF-L, USF-S/ USF-L
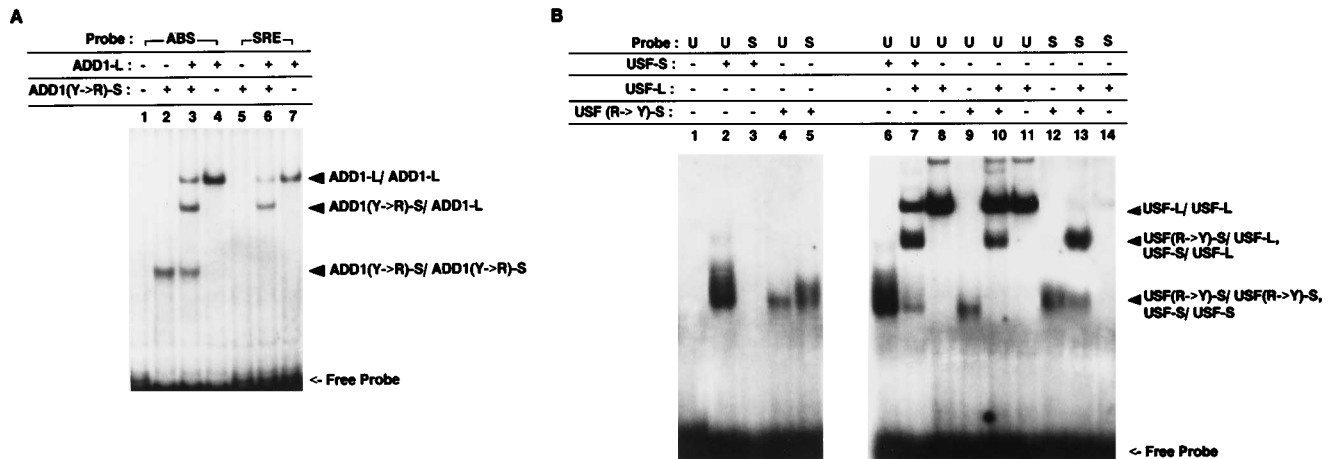◄ USF(R->Y)-S/ USF(R->Y)-S, USF-S/ USF-S
<- Free Probe

FIG. 4. Role of the tyrosine residue in the basic domain of the bHLH-LZ proteins in DNA binding specificity. (A) Changing of DNA binding specificity by the arginine substitution in the basic domain of ADD1/SREBP1. ADD1 (Y-320→R)-S and wild-type protein, ADD1-L, were mixed in the indicated combination and analyzed by EMSA for binding affinity to equal amounts of ABS and SRE-1 probes. (B) Effects of the tyrosine mutation in the basic domain of USF protein. The USF (R-211→Y) mutant was constructed with substitution of an arginine by a tyrosine residue equivalent to that in the wild-type ADD1/SREBP1 protein. The wild-type and mutant USF proteins (R-211→Y) were mixed in the indicated combinations and analyzed by EMSA for binding to equal amounts of USE (U) and SRE-1 (S) probes. Each specific protein-DNA complex is marked by an arrowhead.

might play in the expression of genes linked to adipogenesis, we surveyed the nucleotide database and examined sequences that have been reported to function in genes activated in fat cells. One potential target is the carbohydrate response element (CTCACGTGGT), which appears to be important in the activation of genes involved in lipid metabolism in response to high levels of carbohydrates (31, 32). This sequence, very close to the preferred ADD1/SREBP1 E box, is found in rat S_{14} and fatty acid synthetase genes. Like the ADD1/SREBP1 gene, these genes are expressed at high levels in liver and fat (25, 31, 32). To examine the DNA binding of ADD1/SREBP1 to the carbohydrate response element of the S_{14} gene, we performed an EMSA. As shown in Fig. 5, wild-type ADD1/SREBP1 protein bound the carbohydrate response element of the S_{14} gene

as well as ABS and SRE-1 (lanes 2 to 4). In accordance with data presented above, ADD1(Y-320→R)-S mutant protein bound to the carbohydrate response element of the S_{14} gene (lanes 5 and 7), but did not bind to SRE-1 (lane 6). This result indicated that ADD1/SREBP1 protein bound with high affinity to the carbohydrate response element of the S_{14} gene.

To examine the ability of ADD1/SREBP1 to function as a transcriptional regulator of the S_{14} gene carbohydrate response element, a transient transfection assay was used. As shown in Fig. 6, cotransfection of the wild-type ADD1/SREBP1 expression vector stimulated transcription of a CAT reporter gene linked to multimerized oligonucleotides for the ABS E-box sequence, the SRE-1 sequence, or the S_{14} carbohydrate response element sequence (see Materials and Methods). By contrast, cotransfection of the ADD1/SREBP1 (Y-320→R) expression vector had no effect on the CAT vector containing the SRE-1 sequence (Fig. 6, lane 6) but still stimulated transcription of the ABS CAT and S_{14} CAT constructs (Fig. 6, lanes 3 and 9). These results clearly indicate that ADD1/SREBP1 can function as a transcriptional activator through both E-box and SRE-1 sequences. The carbohydrate

Protein: -   ADD1-S   ADD1(Y->R)-S
Probe: A   A   S   S14   A   S   S14
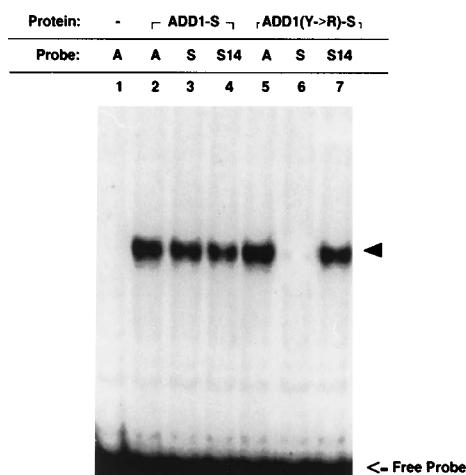   1   2   3   4   5   6   7

◄
<- Free Probe

FIG. 5. DNA binding ability of wild-type or mutant (Y-320→R) ADD1/SREBP1 to the carbohydrate response element of the S_{14} gene. The radiolabeled probes were prepared from oligonucleotides containing the E-box consensus (ABS) sequence, the non-E-box consensus (SRE-1) sequence, or the carbohydrate response element of the S_{14} gene and incubated with bacterially produced ADD1-S or ADD1 (Y-320→R)-S protein. A indicates the ABS probe, S indicates the SRE-1 probe, and S14 indicates the carbohydrate response element E box of the S_{14} gene. The protein-DNA complex is marked with an arrowhead.

ADD1:   -   +   -   -   +   -   -   +   -
ADD1(Y->R): -   -   +   -   -   +   -   -   +
            1   2   3   4   5   6   7   8   9
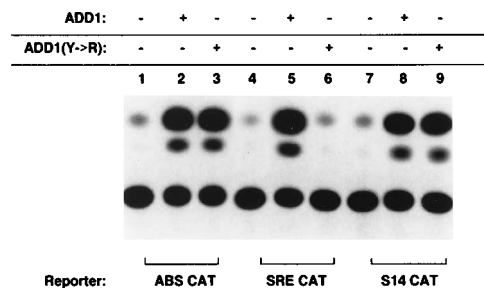
Reporter: ABS CAT   SRE CAT   S14 CAT

FIG. 6. Transcriptional activation by ADD1/SREBP1 through the carbohydrate response element (E-box) and the SRE-1 (non-E-box) target sites. NIH 3T3 cells were transiently cotransfected with wild-type or 320-Y→R mutated ADD1/SREBP1 expression vector with the indicated CAT reporter constructs and assayed for CAT activity (33). ABS CAT, SRE CAT, and S_{14} CAT constructs contain multimerized oligonucleotides for the ABS sequence, SRE-1 sequence, and carbohydrate response element of the S_{14} gene, respectively (see Materials and Methods).

response element of the S$_{14}$ gene is a potential biological target of ADD1/SREBP1.

DISCUSSION

ADD1/SREBP1 is unique among bHLH proteins in that it binds to both E-box and non-E-box DNA sites. Among the E-box sequences, ADD1 strongly prefers a palindrome consisting of two CAC half-sites in the ABS sequence (AT<u>CACG TG</u>A; Fig. 1A). The DNA binding specificity of ADD1/SREBP1 for the core E-box (ABS) CA<u>CGTG</u> versus CA<u>GC TG</u> sequence can be explained by the conserved arginine residue (R-321) immediately next to the tyrosine residue (Y-320) in the basic region of ADD1/SREBP1. It has been shown that the presence of this conserved arginine residue in the basic domain (E-K/R-X-R-<u>R</u>) of proteins such as c-Myc, USF, Max, TFE3 CBF1, and PHO4 is important in binding to the CA<u>CGTG</u> sequence motif versus CA<u>GCTG</u> sequences (12). Although ADD1/SREBP1 binds to the core E-box (CACGTG) site similarly to other c-Myc-related bHLH proteins, these proteins are different in tissue specificity from ADD1/SREBP1, whose mRNA expression is abundant in fat and liver tissue, and they have different flanking sequence preferences (4, 6, 33). Therefore, the various bHLH proteins may regulate gene expression with different tissue and DNA binding specificities despite having the same core E-box binding activity in vitro.

ADD1/SREBP1 also binds to the SRE-1 site (A<u>TCACC CCAC</u>), which is a composite of TCAC and GGGT half-sites. The 5′ TCAC half of SRE-1 is equivalent to an E-box half-site, and it is likely that one subunit of the ADD1/SREBP1 dimer contacts this sequence similarly to other bHLH proteins on their cognate sites. The other subunit of ADD1/SREBP1 binds to the GGGT half of the SRE-1 sequence, and the specificity of this interaction is shown by the invariant sequence of the SRE-1 sites selected by SAAB analysis (Fig. 1B). The dual DNA binding specificity of ADD1/SREBP1 can therefore be described as the alternative binding of one subunit to either TCAC (E-box site) or GGGT (SRE-1 site), with the other subunit contacting a typical E-box sequence.

A tyrosine residue in the basic region of ADD1 (Y-320) is required for binding to the SRE-1 site. bHLH proteins lacking this tyrosine do not bind the SRE-1 site, and mutation of the ADD1/SREBP1 tyrosine to arginine selectively abolishes the binding activity on the SRE-1 site without affecting E-box binding. The role of the equivalent arginine residue in other bHLH proteins has been elucidated by point mutation studies and analyses of crystal structures of several bHLH domains complexed with DNA. For example, in TFEB, this arginine is one of the critical conserved amino acids required to bind DNA (4, 19). Crystal structures of four bHLH protein-DNA complexes all show arginine at this position simultaneously contacting the DNA backbone and a conserved glutamic acid that specifies two bases of the E-box sequence (15, 17, 18, 28). Arginine stabilizes the orientation of the glutamic acid by forming a bridge between it and a DNA phosphate. The tyrosine at this position of the ADD1/SREBP1 basic region cannot make this interaction with glutamic acid. The result of modeling a tyrosine substitution at this position of the Max, USF, MyoD, and E47 crystal structures indicates that the bulky tyrosine ring would crowd the DNA backbone, requiring some adjustment in the DNA and/or protein. Loss of the stabilizing interaction between this residue and the conserved glutamic acid may permit additional local rearrangements at the protein-DNA interface, and these may underlie the observed dual DNA binding specificity of ADD1/SREBP1.

ADD1/SREBP1 is involved in the regulation of several genes of cholesterol homeostasis. For example, this protein has been shown to positively modulate the expression of the low-density lipoprotein (LDL) receptor promoter (37). Cholesterol appears to suppress a proteolytic cleavage that increases the activity of this factor, and thus ADD1/SREBP1 function is sensitive to cholesterol levels (35). All of the known activity of this protein in cholesterol homeostasis is mediated by SRE-1 binding. It is possible that the induction of ADD1/SREBP1 during adipogenesis reflects a different biological context (33). Although fat cells are active in cholesterol metabolism, they play a primary role in the synthesis and breakdown of triglycerides. Adipocytes are active in hydrolyzing triglyceride-rich lipoproteins but are also very efficient in synthesizing lipid from carbohydrates such as glucose, a process which requires the expression of a particular set of genes. The data presented here indicate that ADD1/SREBP1 may play an important role in the latter process through binding to a sequence known as the carbohydrate response element. This element is a classical E-box motif that closely matches one of the ADD1/SREBP1 consensus binding sequences. As expected from the protein binding data shown in Fig. 4 and 5, ADD1/SREBP1 can activate transcription through this element even in the arginine-containing mutation (Fig. 6). Of course, it is also possible that genes involved in other aspects of fatty acid and triglyceride metabolism are also targets for regulation by ADD1/SREBP1 through either E-box or non-E-box binding.

Aside from the unusual structural aspects of the dual binding specificity of ADD1/SREBP1, it is interesting to consider why a single transcription factor possessing this property could be linked to distinct components of lipid metabolism. It is important to note that while cholesterol and triglycerides are synthesized by different biochemical pathways, the fed state in animals usually results in high levels of lipoproteins that carry both lipids. The synthesis of very low density lipoproteins by the liver, for example, requires the synthesis of both cholesterol and triglyceride, and linking both of these pathways to a single transcription factor could be a simple mechanism to accomplish this. The role of ADD1/SREBP1 in various aspects of energy and lipid metabolism must be further explored through genetic experiments in an in vivo context.

REFERENCES

1. **Anthony, C. S. J., P. A. Benfield, R. Fairman, Z. R. Wasserman, S. L. Brenner, W. F. Stafford, C. Altenbach, W. L. Hubbell, and W. F. DeGrado.** 1992. Molecular characterization of helix-loop-helix peptides. Science **255**:979–983.
2. **Beckmann, H., L. K. Su, and T. Kadesch.** 1990. TFE3: a helix-loop-helix protein that activates transcription through the immunoglobulin enhancer μE3 motif. Genes Dev. **4**:167–179.
3. **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.
4. **Blackwell, T. K., J. Huang, A. Ma, L. Kretzner, F. W. Alt, R. N. Eisenman, and H. Weintraub.** 1993. Binding of Myc proteins to canonical and noncanonical DNA sequences. Mol. Cell. Biol. **13**:5216–5224.
5. **Blackwell, T. K., L. Kretzner, E. M. Blackwood, R. N. Eisenman, and H. Weintraub.** 1990 Sequence-specific DNA binding by the c-Myc protein. Science **250**:1149–1151.
6. **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.
7. **Blackwood, E. M., and R. N. Eisenman.** 1991. Max: a helix-loop-helix zipper protein that forms a sequence-specific DNA-binding complex with Myc. Science **251**:1211–1217.

8. **Brennan, T. J., and E. N. Olson.** 1990. Myogenin resides in the nucleus and acquires high affinity of a conserved enhancer element on heterodimerization. Genes Dev. **4:**582–595.

9. **Cabrera, C. V., and M. C. Alonso.** 1991. Transcriptional activation by heterodimers of the achaete-scute and daughterless gene products of Drosophila. EMBO J. **10:**2965–2973.

10. **Carr, C. S., and P. A. Sharp.** 1990. A helix-loop-helix protein related to immunoglobulin E box-binding proteins. Mol. Cell. Biol. **10:**4384–4388.

11. **Caudy, M., H. Vassin, M. Brand, R. Tuma, L. Y. Jan, and Y. N. Jan.** 1988. daughterless, a Drosophila gene essential for both neurogenesis and sex determination, has sequence similarities to myc and the achaete-scute complex. Cell **55:**1061–1067.

12. **Dang, C. V., C. Dolde, M. L. Gillison, and G. J. Kato.** 1992. Discrimination between related DNA sites by a single amino acid residue of Myc-related basic-helix-loop-helix proteins. Proc. Natl. Acad. Sci. USA **89:**599–602.

13. **Davis, R. L., P. F. Cheng, A. B. Lassar, and H. Weintraub.** 1990. The MyoD DNA binding domain contains a recognition code for muscle-specific gene activation. Cell **60:**733–746.

14. **Davis, R. L., H. Weintraub, and A. B. Lassar.** 1987. Expression of a single transfected cDNA converts fibroblasts to myoblasts. Cell **51:**987–1000.

15. **Ellenberger, T., D. Fass, M. Arnaud, and S. C. Harrison.** 1994. Crystal structure of transcription factor E47: E-box recognition by a basic region helix-loop-helix dimer. Genes Dev. **8:**970–980.

16. **Ellis, H., D. R. Spann, and J. W. Posakony.** 1990. Extramacrochaetae, a negative regulator of sensory organ development in Drosophila, defines a new class of helix-loop-helix proteins. Cell **61:**27–38.

17. **Ferré-D'Amaré, A. R., P. Pognonec, R. G. Roeder, and S. K. Burley.** 1994. Structure and function of the b/HLH/Z domain of USF. EMBO J. **13:**180–189.

18. **Ferré-D'Amaré, A. R., G. C. Prendergast, E. B. Ziff, and S. K. Burley.** 1993. Recognition by Max of its cognate DNA through a dimeric b/HLH/Z domain. Nature (London) **363:**38–45.

19. **Fisher, D. E., L. A. Parent, and P. A. Sharp.** 1993. High affinity DNA-binding Myc analogs: recognition by an α helix. Cell **72:**467–476.

20. **Goldstein, J. L., and M. S. Brown.** 1990. Regulation of the mevalonate pathway. Nature (London) **343:**425–430.

21. **Gregor, P. D., M. Sawadogo, and R. G. Roeder.** 1990. The adenovirus major late transcription factor USF is a member of the helix-loop-helix group of regulatory proteins and binds to DNA as dimer. Genes Dev. **4:**1730–1740.

22. **Henthrorn, P., M. Kiledjian, and T. Kadesch.** 1990. Two distinct transcription factors that bind the immunoglobulin enhancer μE5/κE2 motif. Science **247:**467–470.

23. **Hsu, H. L., L. Huang, J. T. Tsan, W. Funk, W. E. Wright, J. S. Hu, R. E. Kingston, and R. Baer.** 1994. Preferred sequences for DNA recognition by the TAL1 helix-loop-helix proteins. Mol. Cell. Biol. **14:**1256–1265.

24. **Hu, Y. F., B. Luscher, A. Admon, N. Mermod, and R. Tjian.** 1990. Transcription factor AP-4 contains multiple dimerization domains that regulate dimer specificity. Genes Dev. **4:**1741–1752.

25. **Jump, D. B., and J. H. Oppenheimer.** 1985. High basal expression and 3,5,3′-triiodothyronine regulation of messenger ribonucleic acid S_{14} in lipogenic tissues. Endocrinology **117:**2259–2266.

26. **Kunsch, C., S. M. Ruben, and C. A. Rosen.** 1992. Selection of optimal κB/Rel DNA-binding motifs: interaction of both subunits of NF-κB with DNA is required for transcriptional activation. Mol. Cell. Biol. **12:**4412–4421.

27. **Lassar, A. B., J. N. Buskin, D. Lockshon, R. L. Davis, S. Apone, S. D. Hauschka, and H. Weintraub.** 1989. MyoD is a sequence-specific DNA binding protein requiring a region of myc homology to bind to the muscle creatine kinase enhancer. Cell **58:**823–831.

28. **Ma, P. C., M. A. Rould, H. Weintraub, and C. O. Pabo.** 1994. Crystal structure of MyoD bHLH domain-DNA complex: perspectives on DNA recognition and implications for transcriptional activation. Cell **77:**451–459.

29. **Murre, C., P. S. McCaw, and D. Baltimore.** 1989. A new DNA binding and dimerization motif in immunoglobulin enhancer binding, daughterless, MyoD, and myc proteins. Cell **56:**777–783.

30. **Murre, C., P. S. McCaw, H. Vaessin, M. Caudy, L. T. Jan, Y. N. Jan, C. V. Cabrera, J. N. Buskin, S. D. Hauschka, and A. B. Lassar.** 1989. Interactions between heterologous helix-loop-helix proteins generate complexes that bind specifically to a common DNA sequence. Cell **58:**537–544.

31. **Shih, H. M., and H. C. Towle.** 1992. Definition of the carbohydrate response element of the rat S_{14} gene. J. Biol. Chem. **267:**13222–13228.

32. **Shih, H. M., and H. C. Towle.** 1994. Definition of the carbohydrate response element of the rat S_{14} gene. J. Biol. Chem. **269:**9380–9387.

33. **Tontonoz, P., J. B. Kim, G. A. Reed, and B. M. Spiegelman.** 1993. ADD1: a novel helix-loop-helix transcription factor associated with adipocyte determination and differentiation. Mol. Cell. Biol. **13:**4753–4759.

34. **Voronova, A., and D. Baltimore.** 1990. Mutations that disrupt DNA binding and dimer formation in the E47 helix-loop-helix protein map to distinct domains. Proc. Natl. Acad. Sci. USA **87:**4722–4726.

35. **Wang, X., R. Sato, M. S. Brown, X. Hua, and J. L. Goldstein.** 1994. SREBP-1, a membrane-bound transcription factor released by sterol-regulated proteolysis. Cell **77:**53–62.

36. **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.

37. **Yokoyama, C., X. Wang, M. R. Briggs, A. Admon, J. Wu, X. Hua, J. L. Goldstein, and M. S. Brown.** 1993. SREBP-1, a basic-helix-loop-helix-leucine zipper protein that controls transcription of the low density lipoprotein receptor gene. Cell **75:**187–197.