MOLECULAR AND CELLULAR BIOLOGY, Feb. 1994, p. 1256–1265
0270-7306/94/$04.00+0

Vol. 14, No. 2

Preferred Sequences for DNA Recognition by the TAL1 Helix-Loop-Helix Proteins

HAI-LING HSU,[1] LAN HUANG,[1] JULIA TSOU TSAN,[1] WALTER FUNK,[2] WOODRING E. WRIGHT,[2] JING-SHAN HU,[3] ROBERT E. KINGSTON,[3] AND RICHARD BAER[1]*

Departments of Microbiology[1] and Cell Biology,[2] University of Texas Southwestern Medical Center, Dallas, Texas 75235, and Department of Molecular Biology, Massachusetts General Hospital, Boston, Massachusetts 02114[3]

Received 30 July 1993/Returned for modification 6 October 1993/Accepted 5 November 1993

Tumor-specific activation of the *TAL1* gene is the most common genetic alteration seen in patients with T-cell acute lymphoblastic leukemia. The *TAL1* gene products contain the basic helix-loop-helix (bHLH) domain, a protein dimerization and DNA-binding motif common to several known transcription factors. A binding-site selection procedure has now been used to evaluate the DNA recognition properties of TAL1. These studies demonstrate that TAL1 polypeptides do not have intrinsic DNA-binding activity, presumably because of their inability to form bHLH homodimers. However, TAL1 readily interacts with any of the known class A bHLH proteins (E12, E47, E2-2, and HEB) to form heterodimers that bind DNA in a sequence-specific manner. The TAL1 heterodimers preferentially recognize a subset of E-box elements (CANNTG) that can be represented by the consensus sequence AACAGATGGT. This consensus is composed of half-sites for recognition by the participating class A bHLH polypeptide (AACAG) and the TAL1 polypeptide (ATGGT). TAL1 heterodimers with DNA-binding activity are readily detected in nuclear extracts of Jurkat, a leukemic cell line derived from a patient with T-cell acute lymphoblastic leukemia. Hence, TAL1 is likely to bind and regulate the transcription of a unique subset of subordinate target genes, some of which may mediate the malignant function of TAL1 during T-cell leukemogenesis.

Several genes have been implicated in the pathogenesis of T-cell acute lymphoblastic leukemia (T-ALL) (39). One of these, the *TAL1* gene (also called *TCL5* or *SCL*), is rearranged in nearly 25% of T-ALL patients, and thus it is likely to be a critical factor in T-cell leukemogenesis (1, 3–6, 9, 10, 15). *TAL1* encodes at least two phosphoproteins, the full-length gene product, $pp42^{TAL1}$ (amino acid residues 1 to 331), and a truncated polypeptide, $pp22^{TAL1}$ (residues 176 to 331) (11). Both contain the basic helix-loop-helix (bHLH) motif, a DNA-binding and protein dimerization domain common to several known transcriptional regulatory factors (for reviews, see references 16, 26, 34, 35, 38, and 44). Although more than 60 different bHLH proteins have been identified to date, the bHLH domain of TAL1 is most related to those encoded by *LYL1* and *TAL2*, distinct genes that are also activated by chromosomal rearrangement in T-ALL (31, 46). Hence, TAL1, TAL2, and LYL1 constitute a discrete subgroup of bHLH proteins implicated in human T-cell leukemia (46).

If TAL1 functions as a transcriptional regulatory factor, then it may promote leukemogenesis by controlling the expression of critical subordinate genes. Protein dimerization appears to be required for DNA recognition by bHLH proteins (34, 35). Thus, the subordinate genes subject to regulation by TAL1 are likely to be determined, at least in part, by its dimerization properties and its DNA-binding preferences. We had previously shown that the bHLH domain of TAL1 mediates protein dimerization with E12 and E47, two bHLH polypeptides encoded by the *E2A* gene (21, 34, 36). The resultant heterodimers (TAL1/E12 and TAL1/E47) bind in a sequence-specific manner to E-box motifs (CANNTG) from the transcriptional enhancer of the immunoglobulin heavy-chain gene (23). However, the immunoglobulin enhancer is unlikely to be a relevant target of TAL1 transcriptional control during either normal or leukemic development. As a first step toward identifying subordinate genes regulated by TAL1 in vivo, we have defined the preferred sites for DNA recognition by TAL1 heterodimers. These results show that TAL1 binds with high affinity to the E-box sequence upon dimerization with any of the known class A bHLH proteins, including the *E2A* gene products, E12 and E47. The TAL1 heterodimers preferentially bind a subset of E-box elements that can be represented by the consensus sequence AACAGATGGT. Thus, TAL1 heterodimers are likely to recognize, and presumably regulate the expression of, a unique subset of subordinate target genes.

MATERIALS AND METHODS

In vitro translation of bHLH polypeptides. Linearized plasmid DNAs were used as templates for in vitro transcription. The plasmid encoding E47S, an 85-residue polypeptide containing the E47 bHLH domain, has been described by Murre et al. (34). Plasmids encoding tag-E47S (E47-tag/pTM4034), tag-TAL1β (talM3/pBS-HA), and tag-TAL1α (talM1/pTM3326) were constructed by inserting the appropriate cDNA fragments into modified cloning vectors; these vectors specify amino-terminal residues with the influenza virus hemagglutinin epitope (YPYDVPDYA) recognized by monoclonal antibody 12CA5 (14). Plasmids encoding the E2-2 (pT7βE2/2) and HEB (pBSATG1-1) polypeptides have been described elsewhere (21, 24). RNA was synthesized from linearized plasmid DNA templates (2 μg) in 100-μl reactions (2 h at 37°C), using T3 or T7 RNA polymerase in

* Corresponding author. Mailing address: Department of Microbiology, UT Southwestern Medical Center, 5323 Harry Hines Blvd., Dallas, TX 75235.

the presence of RNasin (Promega). In vitro translation of RNA was conducted in rabbit reticulocyte lysates as instructed by the manufacturer (Promega). Each in vitro translation reaction was performed in duplicate, that is, in the presence and absence of [^{35}S]methionine. The labeled reactions were then analyzed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis to ascertain both the quality and quantity of in vitro-translated polypeptides. The E2-2 reactions invariably generated smaller products in addition to the intact polypeptide. The smaller products are probably the result of translation from downstream initiation codons, and they are presumably responsible for the faster-migrating electrophoretic mobility shift assay (EMSA) complexes that appear in lanes 4 and 11 of Fig. 5.

CASTing procedure. The target oligonucleotide for CASTing (cyclic amplification and selection of targets) contains PCR primer sites that flank a central core of 35 degenerate nucleotides (17). Five micrograms of the oligonucleotide was converted to double-stranded DNA by annealing an excess of the 3′ primer and performing one primer extension reaction with *Taq* DNA polymerase (17, 18, 45). For the first round of CASTing, a 20-μl binding reaction mixture containing the double-stranded oligonucleotide (10 μl), 8 μl of rabbit reticulocyte lysate (see below), and 2 μl of binding buffer (20 mM *N*-2-hydroxyethylpiperazine-*N*′-2-ethanesulfonic acid [HEPES; pH 7.6], 100 mM potassium chloride, 20% glycerol, 0.2 mM EDTA, 1 mM dithiothreitol) was incubated at room temperature for 20 min. After addition of 2 μl of magnetic beads coated with monoclonal antibody 12CA5, the mixture was incubated for an additional 30 min at 4°C with continuous agitation. The beads were then resuspended in 500 μl of washing buffer (1× phosphate-buffered saline with 0.1% bovine serum albumin and Nonidet P-40), retrieved with a 12-lb (ca. 5.4-kg) magnet, washed four times with 500 μl of washing buffer, and finally resuspended in 100 μl of PCR solution (*Taq*I buffer containing 200 μM each deoxynucleoside triphosphate, and 1 μM concentrations of each of the 3′ and 5′ primers). As described previously, overamplification of the DNA was avoided by monitoring aliquots taken every five PCR cycles by gel electrophoresis (45); amplified DNA from the appropriate sample was precipitated with ethanol, resuspended in a 20-μl binding reaction mixture, and subjected to a new round of CASTing. After six rounds, the selected DNA was PCR amplified and cloned into the M13mp18 phage vector. Individual clones were then subjected to dideoxynucleotide sequence analysis (41).

For CASTing experiments performed with a single in vitro-translated polypeptide, the binding reaction mixtures contained 4 μl of the programmed reticulocyte lysate and 4 μl of unprogrammed lysate. Simultaneous CASTing with two distinct in vitro-translated polypeptides was conducted by combining 4 μl of each programmed lysate. For CASTing experiments with extracts from the Jurkat T-ALL line, each binding reaction mixture included 4 μl of a programmed reticulocyte lysate and 4 μl of a total Jurkat cell extract; total cell extracts were prepared as described by Lassar et al. (29). 12CA5-coated magnetic beads were prepared by mixing 200 μl of M-450 sheep anti-mouse immunoglobulin G Dynabeads (Dynabeads Research Products, Great Neck, N.Y.) with 10 ml of 12CA5 hybridoma cell supernatant; after mixing at room temperature for 16 h, the coated beads were collected with a 12-lb magnet, resuspended in 0.5 ml of washing buffer, recovered again with a 12-lb magnet, and finally resuspended in 200 μl of washing buffer (17).

EMSAs. Nuclear extracts of cultured cells were prepared (30) and used for EMSAs as described previously (23). Each EMSA reaction mixture contained 25 μg of nuclear extract. Some reaction mixtures were supplemented with 1 μl of an appropriate rabbit serum (Fig. 8 and 9). EMSAs with in vitro-translated polypeptides were performed as described previously (23). Western blot (immunoblot) analysis was performed with the anti-E2A or the anti-TAL1 (antiserum 1080) rabbit serum by enhanced chemiluminescence according to the manufacturer's protocol (Amersham International, Amersham, United Kingdom).

Rabbit antisera. Plasmid E12(217-371)/pGEX-KG was used for bacterial expression of GST-E2A(217-371), a glutathione *S*-transferase (GST) fusion protein that includes amino acid residues 217 to 371 of E2A (numbering system of Kamps et al. [27]); this plasmid was constructed by transferring a 0.5-kb *Nco*I-*Xho*I fragment from plasmid pE12R (35) into the pGEX-KG expression vector (20). GST-E2A(217-371) polypeptides were then expressed in *Escherichia coli* and purified by affinity chromatography on glutathione-agarose beads (42). The purified fusion protein was used to produce polyclonal antisera from two rabbits as described previously (11). Two different TAL1-specific rabbit antisera were used in this study. Antiserum 1080 (used in Fig. 9) was obtained by immunizing with GST-TAL1(1-121), a GST fusion protein containing the amino-terminal 121 residues of TAL1. Antiserum 370 (used in Fig. 9 and 10) was raised against GST-TAL1(238-331), a fusion protein containing the carboxy-terminal 94 residues of TAL1 (11).

RESULTS

Preferred sequences for DNA recognition by TAL1/E2A heterodimers. We previously showed that TAL1 interacts in vitro with the E2A proteins (E12 and E47) and that the resulting heterodimers bind with moderate affinity to E-box motifs within the transcriptional enhancer of the immunoglobulin heavy-chain locus (23). It remains to be established whether the leukemic function of TAL1 is mediated by TAL1/E2A complexes present in T-ALL cells. Nevertheless, it seems reasonable to propose that these complexes promote T-ALL by serving as transcriptional regulatory factors. If so, it becomes imperative to identify subordinate genes whose transcription is controlled in T-ALL cells by TAL1. A first step in this process would be to identify the sequence preferences for DNA recognition by TAL1/E2A complexes. Therefore, we used a CASTing procedure to screen pools of degenerate oligonucleotides for high-affinity TAL1 binding sites (17, 18, 45). In vitro translation in reticulocyte lysates was used to produce the truncated *TAL1* gene product (pp22^{TAL1}; residues 176 to 331) with a 15-amino-acid tag (MYPYDVPDYAMGIPI) appended to its amino terminus. The amino-terminal tag includes a short epitope (YPYDVPDYA) recognized by monoclonal antibody 12CA5 (14). Lysates containing this polypeptide (tag-TAL1β) were incubated with an excess of double-stranded oligonucleotides; as described previously, these oligonucleotides contain a 35-nucleotide core of complete sequence degeneracy flanked by common priming sites for PCR (17). DNA-protein complexes were then purified from the mixture by absorption to magnetic beads coated with the epitope-specific monoclonal antibody 12CA5. Bound DNA was released from the beads, amplified by PCR, and subject to another round of CASTing. After six CASTing cycles, the final PCR product was ligated into an M13 cloning vector and the DNA sequences of individual clones were determined.

As shown in Fig. 1A, clones generated by CASTing with

A

1. TAACTGACTTGGCAATCAGCCAGGTTATAGTGCTG
2. ACCACCGCGAAAGTTAGTAGTTTTACGCGCACCCA
3. CGCAGAAGTCTACAGATGCCACGGGTACTGCCCCG
4. GTCGCGGATGGTCGAAATGAGTCAGGTTAAGGCTG
5. ATCATACGCGTCGCACCGCTGCCGCTTCGGGGTCT
6. GCGAACTCTCGGAATCCAGGCTGCAACAGGGCTAA
7. GGATAGTGAAACGCCTGTTGGTTCCGCACAAAGCT
8. TGTTATACCCAGGGTCTTTGCAAGGGCACGGTCCC
9. TGTGCTGGAACAGTGCCAAATTCGGCGCGGTGTTG
10. ACTTACGGTCCGCCCAATGCCACTGTCTGCTACGT
11. ACTTCTGTTGGCTCGCAGCCAAGGCGCGGTTATGG
12. GCGTTCTCGGGTTCAACGTGGGTGACGTACAGCCA
13. CCAGATTGTCCATTTGCTAGGTCATTTACGAAGGG
14. ATTCTGGCTGGGTGCCAGGATGTCATTGTACTGCC
15. AAACCGCGGTATCCTCTTCCTGGTTTCAGCATCGG
16. TTTGACCCACCACGCTGGAACAACAGGGTCCCTCG
17. GGACACTCCGTTTGGTGAGCGTGCAAAAGTCCCTG
18. ATCGACCCTCGCGAACTTGAGTCAGACCTCACCAT
19. TAGGCGAATTCACATTTATCATGCTGGAGATCTTG
20. GTGTCGCATTGGGGTCTGGACCGTAACCAATAGCC

B

1. ACAGATGTTCGCGCACCATCTGGACGCTGAACGAC
2. CCATCTGGCCATTCTGTAATGGACACCATCTGTTC
3. TCCTGAACAGATGGTCGCGACCGGACAAACATCTG
4. AATGTGCAAAACCATCTGTTAGTCCCGACCATCTG
5. ACAGATGTTTCTTTATCATGAGCGAACATCTGGTG
6. AACAGGTGTTCTCGGACCATCTGGACGGATGAGCT
7. CCCGTGTACATCCAGATGGTCCGGTGAACATCTGG
8. TAACAGATGGTAGCCCAGCCATCTGTTGGCGCAGG
9. TCAAACTTGACAGATGTTCATTAAACATCTGTTGC
10. CGCAACATCTGGTCAAGGTGGTTGAACAACACCTG
11. AAACAGATGGCCGACCATCTGTCGAGATCGGCGGG
12. TCCATCTGGTACGGAACATCTGGGCACCATATGGT
13. CAGATGGCACACGATGACCAACTGACCATCTGTTA
14. AACACCTGGACAACAGATGGTACAGGACCAGCTGT
15. CGGAAAACAGATGGTGCAATGCTTCCAGATGGTCG
16. CGAACAGATGTTATGCACCTTGGCATCCAGATGGT
17. GTGAACAGGTGGCTGCCTTATACCATCTGTTCGTG
18. ACAGATGGCAAACTGTACAGCAACAGATGGTCCAG
19. TTGCCTAAACAGATGTTCAGCAGAACATCTGTGCG
20. GCAACACCAATCCAGATGTTTCCTGACCACCTGTG
21. ACATCTGTTTAGTGTACTTGAGACCATCTGGTACC
22. CACCGCTTGATCAAACAGATGGTCGACCATCTGGC
23. CGAACAGATGTTAAGCTCACCAGATGTTTAGGGAC
24. GGAACAGATGTTGGTGAGAGAACATCTGCTGTTCA
25. CGATCAGTAACAGATGTTCCCCCCACCAGATGGCT
26. CCAGACACGACAGATGTTTAGTCCACCATCTGTTT
27. ACAGATGGTGGATGCTCATCTGAACATCTGCTGCG
28. TGCCGCCAACAGATGGCTGGGCTACCATCTGTTAT
29. AACACCTGTTCTCGGACCATCTGGACGGATGAGCT
30. CCAACAGGTGTTATAACCATCTGTTCTACCGGGTG
31. AACAGATGGTACTTTCGCCATCTGTCAACGACTCA
32. GCGAACAGATGGCTCGTAAAACACATTGCGGAGGC
33. GTCAGTTGCACAGATGTTACGAGAACATCTGTTCT
34. TTACGAACATCTGTTCAGTTGCCATCTGTCGTTGC
35. TTACCATCTGGTCAAAACCGCACCATCTGAGAGGG
36. TAACAGATGGTAGCCCAGCCATCTGTTGGCGGCAG
37. AACAGATGGTGTTTAAAACATCTGGTCGCGCGGGC
38. ACCATCTGTTTAGGTCACCATCTGTTCATCGTCGC
39. CACTGGCCAGATGGCAGGAAGAACATCTGCTCCGG
40. CCAGGTGGTGCTCGACCATCTGTTCCAGGTGCATG
41. GAACAGATGGTCTTCTGGACGATAACACCTTGTGT
42. ACAGATGTTCGCGCACGATCTGGACGCTGAACGCA
43. GCGACAGATGGTGAGGTGACCATCTGGGGCACAGC
44. ACAGATGGCAAACTGTACACGAACAGATGGTCCAG

C

E-box core

position	-8	-7	-6	-5	-4	-3	-2	-1	1	2	3	4	5	6	7	8
%C	35	30	23	9	30	100	0	0	1	0	0	3	17	43	23	26
%A	23	15	14	66	61	0	100	0	86	0	0	1	3	20	27	23
%G	28	23	44	14	3	0	0	100	10	0	100	61	1	17	31	32
%T	14	32	19	11	6	0	0	0	3	100	0	35	79	20	19	19
consensus				A	A	C	A	G	A	T	G	G	T			

D

E-box core

position	-8	-7	-6	-5	-4	-3	-2	-1	1	2	3	4	5	6	7	8
%C	10	40	26	8	26	100	0	0	6	0	0	3	17	34	24	39
%A	18	14	12	59	65	0	100	0	86	0	0	1	0	19	34	17
%G	35	17	41	27	9	0	0	95	8	0	100	56	4	35	19	30
%T	37	29	21	6	0	0	0	4	0	100	0	40	79	12	23	14
consensus				A	A	C	A	G	A	T	G	G	T			

FIG. 1. CASTing with TAL1 polypeptides. (A) Nucleotide sequences of 20 individual clones obtained by CASTing with in vitro-translated tag-TAL1β polypeptides. Potential E-box elements (CANNTG) are underlined. (B) Sequences of 44 clones obtained by simultaneous CASTing with in vitro-translated tag-TAL1β and E47 polypeptides. (C) Calculation of a consensus sequence for recognition by tag-TAL1β/E47 heterodimers. The 92 E-box sequences from panel B were aligned from positions −8 to +8. The consensus sequence was determined by calculating the percent frequency of each nucleotide at each position. A frequency of 50% or greater was arbitrarily chosen as a meaningful bias toward a given nucleotide at a particular position. (D) Preferred DNA sequence for recognition by tag-TAL1α/E47 heterodimers. Thirty-five individual clones were selected by simultaneous CASTing with in vitro-translated tag-TAL1α and E47S polypeptides. The 80 E-box elements obtained by CASTing were aligned from positions −8 to +8. The consensus sequence was determined as described above.

tag-TAL1β retained an apparently random sequence within the 35-nucleotide core. Notably, the 20 clones examined contained a total of only two potential E-box motifs (CANNTG), a sum close to that expected for a random sequence of that length. However, the results of the CASTing experiment were dramatically different when the tag-TAL1β lysate was supplemented with a reticulocyte lysate programmed to produce E47S, an 85-residue E2A polypeptide that encompasses the E47 bHLH domain (but does not contain the 12CA5 epitope) (34). Figure 1B shows that each clone selected by CASTing with combined tag-TAL1β-plus-E47S lysates contained two E-box motifs on average. A total of 92 E-box sequences were found within the 44 clones examined. Therefore, the presence of E47S instilled TAL1 with the ability to selectively bind E-box elements, presumably through formation of tag-TAL1β/E47 heterodimers. Moreover, the nucleotides found at the internal and flanking residues of these E boxes were nonrandom and asymmetric, thereby allowing alignment of the 92 E-box elements selected by tag-TAL1β/E47. The consensus binding sequence for tag-TAL1β/E47 calculated in Fig. 1C (AACAGATGGT) reflects clear preferences at both the internal and flanking residues of the E-box motif.

To determine whether the full-length and truncated TAL1 gene products share similar DNA recognition properties, a full-length polypeptide (residues 1 to 331) with an amino-terminal epitope tag (MYPYDVPDYAMGIPI) was also synthesized by in vitro translation. Lysates containing this polypeptide (tag-TAL1α) were then supplemented with the E47S lysate and used for selection through six CASTing cycles. Again, each of the 35 clones analyzed contained an average of two E-box motifs (data not shown). Moreover, alignment of the 80 E-box elements in these clones revealed a consensus binding sequence for tag-TAL1α/E47 identical to that calculated for tag-TAL1β/E47 (Fig. 1D). Thus, the DNA recognition properties of the full-length and truncated *TAL1* gene products appear to be indistinguishable.

To ascertain whether the calculated consensus sequence in fact represents a high-affinity binding site for TAL1 heterodimers, EMSAs were conducted with radiolabeled oligonucleotides containing either the TAL1 consensus E-box sequence or the immunoglobulin μE5 E-box sequence (Fig. 2). These oligonucleotide probes were incubated with reticulocyte lysates programmed to produce either the full-length *TAL1* gene product (tag-TAL1α), the truncated *TAL1* product (tag-TAL1β), or the 85-residue E47S polypeptide.

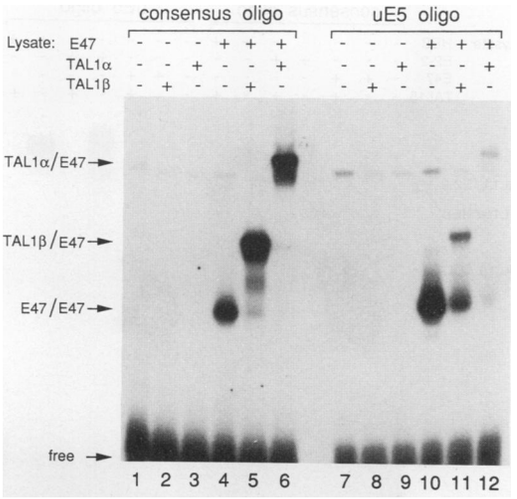

FIG. 2. EMSAs of DNA binding by TAL1/E47 heterodimers. ^{32}P-labeled DNA probes were prepared from oligonucleotides containing the TAL1-binding consensus sequence (top strand, ACCT GAACAGATGGTCGGCT) or the immunoglobulin μE5 sequence (GAACCAGAACACCTGCAGCA). The probes were then incubated with the indicated combinations of in vitro-translated E47S, tag-TAL1α, and tag-TAL1β polypeptides. After electrophoresis on a native 5% polyacrylamide gel, the radiolabeled oligonucleotides were detected by autoradiography. Free oligonucleotides migrated rapidly toward the anode. Protein-DNA complexes with slower mobilities are denoted with arrows, and the presumptive protein composition of each complex is indicated.

As illustrated in Fig. 2, the full-length and truncated TAL1 polypeptides are incapable of binding the TAL1 consensus oligonucleotide, presumably because of ineffectual homodimerization (lanes 2 and 3); this observation is consistent with the results of CASTing experiments performed with TAL1 polypeptides alone (Fig. 1A). Incubation of the same oligonucleotide probe with an E47S-programmed lysate results in the formation of a protein-DNA complex that presumably represents binding by E47/E47 homodimers (Fig. 2, lane 4) (34). However, protein-DNA complexes with distinct electrophoretic mobilities arise when the E47 lysate is supplemented with lysates containing either the truncated or full-length TAL1 polypeptide (lane 5 or 6, respectively). These complexes (designated TAL1β/E47 and TAL1α/E47 in Fig. 2) are clearly heteromeric, since their formation is dependent on the presence of both TAL1 and E47 and is abrogated by preincubation with polyclonal antisera specific for TAL1 (data not shown). The TAL1α/E47 and TAL1β/E47 heterodimers bind less effectively to an oligonucleotide probe containing the μE5 E box from the immunoglobulin heavy-chain gene enhancer (Fig. 2, lanes 11 and 12). It is also apparent that the presence of TAL1 polypeptides reduces complex formation between E47/E47 homodimers and the μE5 probe (compare lanes 11 and 12 with lane 10); this presumably reflects the recruitment of E47 polypeptides into bHLH heterodimers (i.e., TAL1/E47) that bind the μE5 sequence poorly. Clearly, the TAL1 consensus E box derived from the CASTing experiments represents a higher-affinity binding site for TAL1/E2A heterodimers, and thus it may resemble physiologic sites for TAL1 recognition that function in vivo.

DNA recognition by TAL1/E2-2 and TAL1/HEB heterodimers. Recent studies have uncovered two distinct human genes, *E2-2* and *HEB* (also called *HTF4*), that are highly related to *E2A* (21, 24, 47). The products of the three loci, collectively termed E proteins or class A bHLH proteins, exhibit broad, overlapping patterns of expression and share extensive amino acid homology that extends both within and without their bHLH domains. Therefore, the CASTing procedure was used to determine whether TAL1 also binds DNA in association with E2-2 and HEB. A reticulocyte lysate containing the tag-TAL1β polypeptide was prepared and supplemented with lysates programmed to produce E2-2. The in vitro-translated E2-2 polypeptide consisted of the carboxy-terminal 618 residues of human E2-2, including the entire bHLH motif (but did not possess the 12CA5 epitope) (21). The combined tag-TAL1β-plus-E2-2 lysates were then used for selection through six CASTing cycles, and 30 of the selected clones were examined by nucleotide sequence analysis. Again, each of these clones displayed approximately two E-box sequences within the 35-nucleotide core (Fig. 3A). Moreover, alignment of the 59 E-box elements revealed a consensus binding sequence for tag-TAL1β/E2-2 identical to that determined for the TAL1/E2A heterodimers (Fig. 3B). Similar results were also obtained in CASTing experiments with HEB. Thus, tag-TAL1β lysates were supplemented with lysates containing HEBr, a truncated polypeptide composed of the 347 carboxy-terminal residues of human HEB, including the intact bHLH domain (but not the 12CA5 epitope) (24). After six cycles of CASTing with the combined tag-TAL1β-plus-HEBr lysates, 29 of

A

1. CAGATGGCCTGCCACCCGTTTATGCAGCAGATGCT
2. CACCATCTGTTCGCTGCAGTCACCATCTGTTGTCT
3. AACAGGTGTTCTCGGACCATCTGGACGGATGAGCT
4. ACATCTGTTTAGTGTACTTGAGACCATCTGGTACC
5. CCCCCGATTCAAGGAGGTGGCTCAGCATCTGTTCG
6. ACAGATGGTCGCATGACAGATGGTATCAAGCTGTT
7. ACAGATGGTTTGGTCCCTGCCGTCCAGATGGCCGG
8. CGAACAGATGTTATGCACCTTGGCATCCAGATGGT
9. ACAGATGGCAAACTGTACACGAACAGATGGTCCAG
10. CGATCACTAACAGATGTTCCCCCCACCAGATGGCT
11. ACATCTGTTTACTCTACGGCAGACCATCTGGTACC
12. CAAACAGCCATCTGTTCAGAGGAACATCTGTTCAG
13. GACACCTGGACAACAGATGGTACTGGACCAGCTGT
14. TTGCCTAAACAGATGTTCAGCAGAACATCTGGCGA
15. AACAGATGGTGTTTAAAACATCTGTCGCGCGGGCG
16. GCGAACAGATGGCTCGTAAAACACATTGCGGAGGC
17. CTGAACAGATGTTCCTCTGAACAGATGGCTGTTTG
18. GCAGATGGCAAACTGTACGCGAACAGATGGTCCAG
19. CCATCTGGACGATGCCATCTGTTGCCCTCTCGAAG
20. CTACAGATGGTTATGCACAGGAGCGACATCTGTCA
21. GCTCAGAGGGCAACAGATGGCATCGTCAGATGGTC
22. GAACAGATGGCTGAGACCCACCTTGAATCGGGGGT
23. CATCTGTTTAGTGTACTTGAGACCATCTGGTACCG
24. GTGACAGATGTCGCCCCGTGCATAACCATCTGTTG
25. CAGATGGTTGGGAGAGTTGCCAGCTACCATCTGGT
26. ACAGATGGCTCTATACGTATAACAGATGTTCTTTC
27. ACATCTGTTCTGGCCATCTGTTTGCGCGACATATG
28. TACCATCTGTTTCGTGCAGTCACCATCTGTTGTCG
29. GCGAACAGATGTTCATTCACCATCTGTTTCATCAA
30. ACATCTGTTTAGTGTACATGAGACCATCTGGTACC

B

E-box core

position	-8	-7	-6	-5	-4	-3	-2	-1	1	2	3	4	5	6	7	8
%C	32	28	23	0	21	100	0	0	2	0	0	3	30	39	21	35
%A	10	12	19	74	72	0	100	0	95	0	0	0	0	17	25	16
%G	43	35	37	14	5	0	0	98	3	0	100	63	0	20	14	26
%T	15	25	21	12	2	0	0	2	0	100	0	34	70	24	40	23
consensus				A	A	C	A	G	A	T	G	G	T			

FIG. 3. Preferred sequences for DNA binding by TAL1/E2-2 heterodimers. (A) Sequences of 30 individual clones selected by simultaneous CASTing with in vitro-translated tag-TAL1β and E2-2 polypeptides. Potential E-box elements (CANNTG) are underlined. (B) The 59 E-box sequences obtained by CASTing were aligned from positions −8 to +8. A consensus sequence for preferential recognition by TAL1/E2-2 was calculated as described for Fig. 1.

A

1. TACGAACAGATGTTTTGTCCAGATGGTCCGGCTTA
2. CACCATCTGTCACACGTGCCATCTGTTCGGGGTCC
3. AACAACATCTGTTTGCTCTTTCTTACCATCTGCAC
4. CACCATCTGTTTAGCGTCGCCAGATGTTCACATCA
5. ACAGATGTTCGCGCACCATCTGGACGCTGAACGCA
6. CCCAGGAAACATCTGTCGCGACATCTGTTCAATGA
7. GAACATCTGTTCGCACTGCCGGTGACAGGTGCGTG
8. CCATCTGGACAGCTTACCATGCACCATCTGTTTAT
9. ACAACAGATGTTTACAATCTTCGTATGACCATCTG
10. GCCCATACCATCTGTTTTTCGTGACCATCTGTTCG
11. TAACATCTGTTCTGACATTCTCTCAACACCTGTTG
12. CCTGAACAGATGGTCGTGCCGCCATCTGTTCCTTG
13. CAGATGGCGCCTTGGCCATATGGTGGTACCATCTG
14. AACAGATGGCGTACGACCAGATGTTCTCATATCCG
15. ACAGATGTTACATAACAGATGTTTTCCACGGGCTT
16. ACCATCTGTTTGGGTCGTAACAGGCCATCTGTTGC
17. CCATCTGCTAGTCACAATCTGCGCCATCTGTTTAG
18. TCACCAGATGGCGCTGTTAAGGACCGGACCATCTG
19. TCACCAGATGTTGCGGCTACTATACCATCTGATCC
20. ACAGATGTTAGCTCGTGAACCCGTCACATCTGGTC
21. CACCTGCTTGATAGCGCTTGAAAAAACATCTGTTC
22. CAGATGTTCCGAACAGATGGTCACAACATGCGGTT
23. ACAACAGATGGTCCGTAGTTGTCACCAGATGTTAC
24. GCTGGCGCTCTGCCAACTCACCATCTGTTTTCTCT
25. AACATCTGTTCGCCGAGCCATCTGCTTACCCCACG
26. AAACAGATGGTCCTGTGAGTTTTCCAGATGTTCGG
27. CCATCTGTTCAACAACTTAAATACCATCTGTTCGC
28. GCTGGCGCTCTGCCAACTCACCATCTGTTTTCTCT
29. AAAAAACATCTGTTGCTGGTGAGTTCGACCATCTG

B

E-box core

position	-8	-7	-6	-5	-4	-3	-2	-1	1	2	3	4	5	6	7	8
%C	29	32	24	0	22	100	0	0	0	0	0	2	17	31	40	29
%A	18	12	22	75	64	0	100	0	95	0	0	0	0	22	22	14
%G	32	22	43	10	10	0	0	98	5	0	100	58	4	33	18	39
%T	21	34	11	15	4	0	0	2	0	100	0	40	79	14	20	18
consensus				A	A	C	A	G	A	T	G	G	T			

FIG. 4. Preferred sequences for DNA binding by TAL1/HEB heterodimers. (A) Sequences of 29 clones selected by simultaneous CASTing with in vitro-translated tag-TAL1β and HEBr polypeptides. Potential E-box elements are underlined. (B) The 57 E-box sequences obtained by CASTing were aligned from positions −8 to +8. A consensus sequence for preferential recognition by TAL1/HEB was calculated as described for Fig. 1.

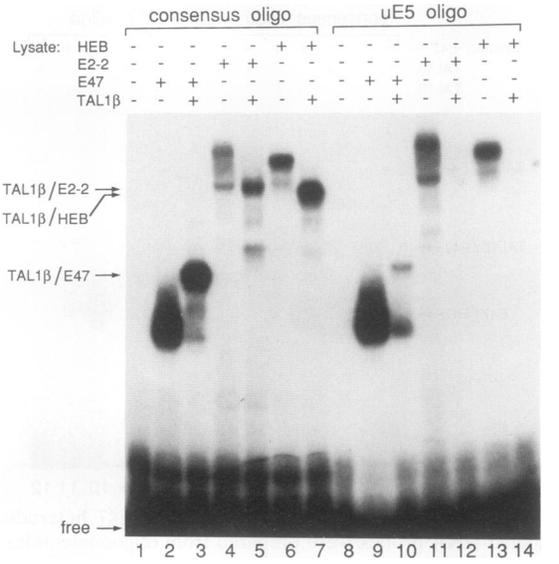

FIG. 5. DNA-binding properties of TAL1/E2-2 and TAL1/HEB heterodimers evaluated by EMSAs. Different combinations of in vitro-translated HEBr, E2-2, E47S, and tag-TAL1β polypeptides were incubated with ^{32}P-labeled oligonucleotide probes representing either the TAL1-binding consensus E box or the immunoglobulin μE5 E box (see Fig. 2). After electrophoresis on a native 5% polyacrylamide gel, the radiolabeled probes were detected by autoradiography. Free oligonucleotide probes migrated rapidly toward the anode. The protein-DNA complexes representing TAL1 heterodimers, TAL1/E47 (lanes 3 and 10), TAL1/E2-2 (lane 5), and TAL1/HEB (lane 7), are denoted with arrows.

the selected clones were subjected to sequence analysis. Each clone contained an average of two E-box sequences (Fig. 4A). As shown in Fig. 4B, alignment of the 57 E-box elements revealed a consensus binding sequence identical to that determined for TAL1/E2A and TAL1/E2-2 heterodimers.

EMSAs were performed to confirm that TAL1 interacts with the E2-2 and HEB polypeptides to form bHLH heterodimers with DNA-binding potential. Therefore, double-stranded oligonucleotide probes containing either the TAL1 consensus E box or the immunoglobulin μE5 E box were prepared. These radiolabeled probes were then incubated with an in vitro-translated polypeptide representing E47, E2-2, or HEB. As illustrated in Fig. 5, each of these polypeptides bound the TAL1 consensus E box, presumably by formation of class A bHLH homodimers (lane 2, E47/E47 homodimer; lane 4, E2-2/E2-2; lane 6, HEB/HEB). However, in the presence of a TAL1 polypeptide (tag-TAL1β), each homodimeric complex was replaced by a new species with an electrophoretic mobility consistent with formation of a TAL1 heterodimer (lane 3, TAL1/E47; lane 5, TAL1/E2-2; lane 7, TAL1/HEB). Each of the presumptive heterodimers, but not the corresponding homodimers, was specifically disrupted by preincubation with TAL1 antisera (data not shown). Moreover, the heterodimers, but not the homodimers, bound far more effectively to the TAL1 consensus E box than to the immunoglobulin μE5 E box (compare lanes 3, 5, and 7 with lanes 10, 12, and 14). In sum, these results indicate that each of the known class A bHLH proteins (E12, E47, E2-2, and HEB) can interact with TAL1 to form bHLH heterodimers with DNA-binding activity. Moreover, heteromeric association with TAL1 appears to be favored, at least under in vitro conditions, over the formation of class A homodimers (Fig. 5).

Asymmetric DNA recognition by E2A homodimers. Previous studies had shown that the DNA-binding site of a bHLH dimer (i.e., the E-box motif) can be viewed as a combination of two half-sites, each of which represents the recognition sequence for one of the two dimerized bHLH polypeptides (7). Experimental strategies related to, but distinct from, the CASTing procedure have been used to determine the half-site recognized by the E2A proteins, E12 and E47 (7, 45). However, the E2A half-site (AACAC) used by MyoD1/E2A and myogenin/E2A heterodimers is not found within the consensus sequence determined for recognition by TAL1/E2A heterodimers (AACAGATGGT; Fig. 1). Therefore, the nature of the E2A half-site was investigated further by using the CASTing procedure. A reticulocyte lysate was programmed to produce tag-E47S, a polypeptide that encompasses 85 residues of E47, including the entire bHLH domain, and harbors an amino-terminal tag (MYPYDVPDY AMAIDID) containing the 12CA5 epitope. After six cycles of CASTing with the tag-E47 lysate, 29 of the selected clones were subjected to nucleotide sequence analysis. Most of the clones contained a single E-box motif (Fig. 6A). Notably, 31 of the 38 E-box elements shared a common central sequence that is asymmetric (CAGGTG). Although it is surprising that E47/E47 homodimers preferentially bind an asymmetric sequence, this phenomenon had already been described by Blackwell and Weintraub (7). In any case, the asymmetry facilitated alignment of the 38 E-box elements and calcula-

A

1. CAAAGGTTATCCTCGTGAACACCTGCAACGTCGGG
2. ACACCTGTTCATCGCGATCGGAATACCAGCCATAC
3. GGTTAAAACCCTGAACACCTGCGATTATCGTTGTG
4. TGCCGTTAATATATCTGCAGGTGGTCACGCAGGGT
5. TCAATTTAATAACGTAAACACCTGGTGGGGGATGT
6. CACGTGCTAGCAATTAACACCTGCGATCATGGGGG
7. GACAGAAACAGGTGTTACCTAGCAAAGGGGGGGCC
8. ATCGCATCTTTTGCAGGTGTTACCTCTCTTCCCTG
9. TCTTGCTATTCCTTGCTGCAACACCTGTTCGCTCC
10. TTCAGCACCTGCAGTTTTTCCCAGGTGGACGGCAG
11. GCGTACTACCCTGACAGATGTTCTCGAACAGGTGC
12. TAACACCTGTTAGTGTTAAGACGTAGCAGCAGGGT
13. TCACTTTCCCAATACTTCGTGATAAGTAACACCTG
14. GTGGTTGTGTTCGGACGTAACAGGTGTTCAGCTAG
15. TAACGCACCTGTTCAGCCCTGAAGTGAGTAGTCGA
16. AGTGCCATGTAATCTTAAGGCCAGGTGTTACCTCG
17. GAAGCCTAACCTAAGTAACACCTGTTCATGGTCCG
18. CAGGTGTACAACAGGCGTTCCGCCCAGGAACAGTG
19. TGAACACCTGCGCACCCTTATTATTGCCCTTGTTA
20. AAGTAAGACCAGGTGCAGTACTCCACAGGTCTAGC
21. GATAAACAGGTGTTCAACAGGTAATACAGCCATCC
22. GGGCGTATATCGGACCACCTGCAGTGCACCTGCTG
23. ACACCTGTTATTATTGAGTACCCGGACCTTCTCTT
24. AAGTGCACCTGCGTCTCAGATGTGCGCAGATGCTT
25. TTCTATCTCACAGATGTGGGAACACCTGTTTTTTC
26. CATATGGCCGCGTACTCAGTCGATTCTAACACCTG
27. AAGCGCGCTTGGCCCAACTTGGGCCTGAACACCTG
28. GCGTATAGATTCAACCCCATCACAGGTGTTACGAG
29. GGGCTTGCACCTGTTTTCTTCTATACTTCCATGTG

B

E-box core

position	-8	-7	-6	-5	-4	-3	-2	-1	1	2	3	4	5	6	7	8
%C	24	24	17	26	14	100	0	5	16	0	0	24	3	46	27	37
%A	41	31	17	50	50	0	100	0	0	0	0	0	20	34	34	25
%G	17	14	30	7	33	0	0	92	84	0	100	13	8	11	24	22
%T	17	31	36	17	3	0	0	3	0	100	0	63	69	9	15	16
consensus				A	A	C	A	G	G	T	G	T	T			

FIG. 6. Preferred sequences for DNA binding by E47 homodimers. (A) Sequences of 29 individual clones selected by CASTing with in vitro-translated tag-E47S polypeptides. Potential E-box elements (CANNTG) are underlined. (B) The 38 E-box elements obtained by CASTing were aligned from positions −8 to +8. A consensus sequence for preferential recognition by E47 homodimers was calculated as described for Fig. 1.

tion of a consensus sequence for DNA recognition by E47 homodimers (AACAGGTGTT) (Fig. 6B). Despite asymmetry at the central residues, the flanking sequences of the consensus E-box motif are clearly symmetric. As discussed below, these data permit the assignment of half-sites for DNA recognition by TAL1 heterodimers.

DNA recognition by TAL1 in association with factors from leukemic T cells. Our results show that the DNA-binding potential of TAL1 is expressed upon interaction with any of the known class A bHLH proteins. Moreover, previous studies indicate that class A proteins are present in at least some leukemic cell lines derived from T-ALL patients (2). Nevertheless, it is possible that TAL1 preferentially forms heterodimers with other, as yet unidentified bHLH factors present in T-ALL cells. To investigate this possibility, CASTing experiments were again performed with reticulocyte lysates that contain the truncated TAL1 polypeptide with an amino-terminal epitope tag (tag-TAL1β). However, in this case, the tag-TAL1β lysate was supplemented with a total cellular extract from the Jurkat T-ALL line. This mixture was used for selection through six CASTing cycles, and 30 of the selected clones were examined by nucleotide sequence analysis. Each clone possessed approximately two E-box motifs within the 35-nucleotide core (Fig. 7A), and a consensus binding sequence was readily calculated upon alignment of the 58 E-box elements (Fig. 7B). Notably, this consensus is identical to the preferred binding sequence for DNA recognition by TAL1 heterodimers involving the

A

1. GTGCCATCTGTTTATAGGGCGAACATCTGTTTGTC
2. CCATCTGTTGGTACACACCATCTGTTCGCCGGCTG
3. ACAGATGTTCCTTAACAGCGGCTCTCGAGCATCTG
4. CAGACGGTATCGTTAACAGATGTTTCGGCCATCTG
5. CAGATGTTACCAGCTCGCTTGTGCAGCCACCTGTT
6. GCCAAGCACCTGGTTGACGGAGAGGACCAGATGGC
7. TCCACCAGATGGTTATTCTTGCAGGCAGATGGCCA
8. CAGATGTTCACCAGATGGTTCGGAAACGTGTGTTT
9. ACATCTGTGTGTACGGCAGGTGTTGTGTTTGGCGC
10. AAACGGCTCTGACCATCTGTTCACTTGCAAGGGGA
11. GCCAAAGGAACATCTGGTTGCGAAAACATCTGTTC
12. AAACAGATGTTTTAAGCGCCATCGGTTTCGCCGGC
13. CAGATGGTCGCCGTTGGTGAAAACCATCTGTCAGC
14. TGACTGGTGGAACAGATGTTTTGGTGGCAGATGTT
15. ACAGATGGTCAAGTGAACTGCCTACTGGGTGCCCC
16. CACCATCTGTTCGCTGCAGTCACCATCTGTTGTCT
17. CCATCTGTCAGGTAAACATCTGCTTCTTGGAGCGG
18. ACAGATGTTCGCGCACCATCTGGTCGCTGAACGCA
19. CTAGGTACCATCTGCTGCTTGTCAAATCAAGGTGT
20. AAACACCTGCTCGCCTAGGTCCACCATCTGTTGAA
21. ACATCTGTTCGCTCGCACCATCTGTCAGCAGCTGT
22. CCATCTGTTAGGATTCCAGGTGTTAACAGTGTGCC
23. GCCCATACCATCTGTTCTTCGTGACCATCTGTTCG
24. GCCAAAGGAACATCTGGTTGCGAAAACATCTGTTC
25. CAGATGTTCCGAACAGATGGTCACAACATGCGGTT
26. GGTATCGTGCGCCATCTGGACCAGTTCGACATCTG
27. AACCGCATGTTGTGACCATCTGTTCGGAACGTCTG
28. ATTACAGATGGCACAAACAGATGTTCGCGTGGGAC
29. AACACCATCTGTCACCCACCCAACCATCTGGAACG
30. CAGACCCTTGTAACATCTGTCGCTGCCGCCATCTG

B

E-box core

position	-8	-7	-6	-5	-4	-3	-2	-1	1	2	3	4	5	6	7	8
%C	14	48	17	2	19	100	0	2	2	0	0	5	16	37	28	33
%A	27	11	26	69	61	0	100	2	88	0	0	2	4	16	25	20
%G	40	23	36	19	16	0	0	96	10	0	100	53	0	23	28	29
%T	19	18	21	10	4	0	0	0	0	100	0	40	80	24	19	18
consensus				A	A	C	A	G	A	T	G	G	T			

C

E-box core

position	-8	-7	-6	-5	-4	-3	-2	-1	1	2	3	4	5	6	7	8
%C	10	40	26	8	26	100	0	0	6	0	0	4	17	34	24	39
%A	18	14	12	59	65	0	100	0	86	0	0	1	0	19	34	17
%G	35	17	41	27	9	0	0	96	8	0	100	56	4	35	19	30
%T	37	29	21	6	0	0	0	4	0	100	0	40	79	12	23	14
consensus				A	A	C	A	G	A	T	G	G	T			

FIG. 7. DNA recognition by TAL1 polypeptides in the presence of leukemic cell extracts. (A) Sequences of 30 individual clones selected by simultaneous CASTing with in vitro-translated tag-TAL1β and total cell extract from the Jurkat T-ALL line. Potential E-box elements (CANNTG) are underlined. (B) The 58 E-box elements obtained by CASTing were aligned from positions −8 to +8. A consensus sequence was calculated as described for Fig. 1. (C) Thirty-four clones were selected by simultaneous CASTing with in vitro-translated tag-TAL1α and total cell extract from the Jurkat T-ALL line. The 51 E-box elements obtained by CASTing were aligned from positions −8 to +8. A consensus sequence was calculated as described for Fig. 1.

known class A proteins (Fig. 1, 3, and 4). The same result was obtained in CASTing experiments with reticulocyte lysates containing full-length TAL1 polypeptides. Thus, tag-TAL1α lysates supplemented with a Jurkat cell extract were used for selection through six cycles of CASTing. Again, most of the 34 selected clones displayed one or two E-box sequences within the 35-nucleotide core (data not shown). Moreover, the consensus sequence derived by alignment of the 51 E-box elements (Fig. 7C) is also identical to the preferred recognition sequence of TAL1 heterodimers involving class A bHLH polypeptides (Fig. 1, 3, and 4). These data strongly imply that the DNA-binding activity of TAL1 in Jurkat T-ALL cells is expressed upon dimerization with known class A proteins or with other highly related bHLH polypeptides.

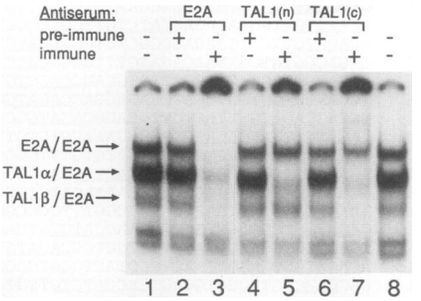

FIG. 8. DNA recognition by the endogenous TAL1/E2A complexes of leukemic T cells. A double-stranded ^{32}P-labeled oligonucleotide probe (upper strand, ACCTGAACAGATGGTCGGCT) was incubated with nuclear extracts from Jurkat cells. For some binding reactions, the extracts were preincubated with specific rabbit antisera (lanes 3, 5, and 7) or the corresponding preimmune sera (lanes 2, 4, and 6). The antisera were raised against either E2A (lanes 2 and 3), the amino-terminal 121 residues of TAL1 (lanes 4 and 5), or the carboxy-terminal 94 residues of TAL1 (lanes 6 and 7). The binding reactions were fractionated by electrophoresis on a native 5% polyacrylamide gel, and the radiolabeled probe was detected by autoradiography. Protein-oligonucleotide complexes representing E2A homodimers (E2A/E2A) or TAL1 heterodimers (TAL1α/E2A and TAL1β/E2A) are denoted with arrows.

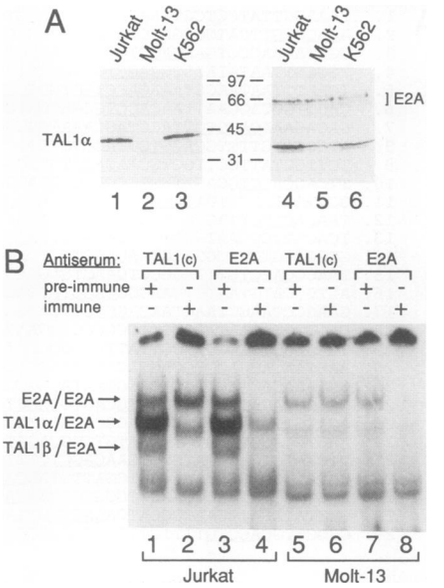

FIG. 9. Correlation between TAL1 expression and formation of presumptive TAL1 heterodimers. (A) Western analysis of total extracts from the indicated cell lines. Immunoblotting was performed with rabbit antisera raised against the amino-terminal 121 residues of TAL1 (lanes 1 to 3) or residues 217 to 371 of E2A (lanes 4 to 6). Sizes are indicated in kilodaltons. (B) Nuclear extracts from the Jurkat (lanes 1 to 4) and Molt-13 (lanes 5 to 7) cell lines were evaluated by an EMSA as described for Fig. 8. Binding reaction mixtures were preincubated with the indicated antisera (lanes 2, 4, 6, and 8) or the corresponding preimmune sera (lanes 1, 3, 5, and 7).

DNA recognition by the endogenous TAL1/E2A complexes of leukemic T cells. To evaluate the DNA-binding activity of endogenous TAL1 polypeptides from leukemic T cells, EMSAs were conducted with a radiolabeled oligonucleotide probe containing the TAL1 consensus E-box sequence. As illustrated in Fig. 8, incubation of this probe with nuclear extracts from Jurkat cells generated several distinct protein-DNA complexes (lanes 1 and 8). Three of these complexes, marked with arrows in Fig. 8, were eliminated by preincubation with an anti-E2A rabbit serum (lane 3) but not with the corresponding preimmune serum (lane 2). One complex (marked TAL1α/E2A) was also abrogated by preincubation with an antiserum raised against the amino-terminal 121 residues of human TAL1 (lanes 4 and 5). Two complexes (marked TAL1α/E2A and TAL1β/E2A) were specifically eliminated by preincubation with an antiserum raised against the carboxy-terminal 94 residues of TAL1 (lanes 6 and 7). The pattern of immune reactivity and the known molecular weights of the E2A (p67), full-length TAL1α (pp42^{TAL1}), and truncated TAL1β (pp22^{TAL1}) polypeptides are consistent with the inferred protein composition ascribed to each complex in Fig. 8. Hence, the complex with the slowest mobility is likely to represent DNA recognition by E2A homodimers (i.e., E12/E12, E47/E47, and E12/E47). In contrast, the other two complexes presumably consist of heterodimers involving both TAL1 and E2A polypeptides (TAL1α/E2A and TAL1β/E2A).

To confirm the protein compositions assigned to these complexes, we performed parallel EMSAs with nuclear extracts from T-ALL cell lines that either do (Jurkat) or do not (Molt-13) express TAL1. Western analysis demonstrated that the two lines harbor equivalent levels of E2A but that TAL1 expression is restricted to Jurkat cells (Fig. 9A). As illustrated in Fig. 9B, EMSAs revealed an E2A/E2A homomeric complex in Molt-13 cells (lane 7) that is specifically abrogated by the E2A antiserum (lane 8). As expected, however, Molt-13 cells do not have complexes corresponding to the TAL1α/E2A and TAL1β/E2A heterodimers. Thus, the presence of TAL1 heteromeric complexes in T-ALL cells correlates with expression of TAL1 polypeptides. This correlation has been extended to other T-ALL lines that either do (RPMI8402) or do not (PEER) express *TAL1* (data not shown).

It may be significant that the E2A antiserum eliminated the formation of most (~90%) but not all of the heteromeric TAL1 complexes in Jurkat cells (Fig. 8, lane 3). Numerous control experiments have shown that the E2A antiserum recognizes E12 and E47 but not other known class A bHLH proteins such as E2-2 and HEB (data not shown). Therefore, as discussed below, some of the protein-DNA complexes that are resistant to disruption by anti-E2A may represent TAL1 heterodimers involving other class A proteins (e.g., TAL1α/HEB).

To evaluate the sequence specificity of DNA recognition by the bHLH complexes from Jurkat cells, EMSAs were performed in the presence of increasing levels of unlabeled oligonucleotide probes. These competitor probes were derived from either wild-type oligonucleotides (ACCTGAA CAGATGGTCGGCT) or corresponding mutant oligonucleotides that bear two nucleotide substitutions in the E-box core (ACCTGAACCGATTGTCGGCT). As shown in Fig. 10, the formation of radiolabeled complexes representing E2A homodimers (E2A/E2A) and TAL1 heterodimers (TAL1α/E2A and TAL1β/E2A) is eliminated by preincubation with an excess of the wild-type competitor (lanes 2 to 4) but not the mutant competitor (lanes 5 to 7). Thus, the bHLH complexes from Jurkat cells bind DNA in a sequence-specific manner that is dependent on the E-box motif.

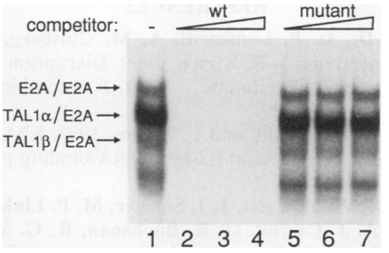

FIG. 10. Sequence specificity of endogenous TAL1/E2A complexes from leukemic T cells. Nuclear extracts from Jurkat cells were evaluated by an EMSA as described for Fig. 8. For some binding reactions, the extracts were preincubated with a 250-fold (lanes 2 and 5), 500-fold (lanes 3 and 6), or 750-fold (lanes 4 and 7) excess of unlabeled oligonucleotides; the competitor oligonucleotides had either the wild-type (wt) TAL1-binding sequence (lanes 2 to 4) or the corresponding mutant sequence (lanes 5 to 7).

DISCUSSION

The protein dimerization and DNA-binding properties of TAL1. Murre et al. suggested that the various bHLH proteins can be divided provisionally into at least three categories (35). These include the broadly expressed class A proteins (E12, E47, E2-2, and HEB), the tissue-specific class B proteins (e.g., myogenic bHLH factors such as MyoD1), and the class C proteins, which feature a tandem arrangement of bHLH and leucine zipper motifs (the bHLH-Zip domain). In at least four respects, the dimerization properties of TAL1 resemble those of known class B proteins. First, TAL1 polypeptides do not form bHLH homodimers with detectable DNA-binding activity, as determined by both the CASTing experiments and EMSAs. Second, the same methods revealed that the DNA-binding potential of TAL1 is expressed upon heteromeric interaction with any of the known class A proteins, including E12, E47, E2-2, and HEB. Third, EMSAs indicate that in the presence of TAL1, class A proteins preferentially bind DNA as heterodimers. Thus, although class A proteins have the potential to form homodimers, these homomeric interactions may be less stable than heterodimer formation with TAL1. Fourth, Sun et al. have shown that TAL1 interacts weakly, if at all, with Id1 and Id2, regulatory proteins that inhibit DNA binding by class A bHLH proteins (43). In sum, these observations indicate that the DNA-binding activity, and presumably the functional properties, of TAL1 are contingent upon heterodimer formation with other bHLH proteins.

We previously showed that TAL1/E2A heterodimers bind with moderate affinity to E-box elements within the transcriptional enhancer of the immunoglobulin heavy-chain gene (23). The CASTing experiments described here were pursued in order to identify high-affinity sites for DNA recognition by TAL1. In the presence of E47, TAL1 polypeptides selectively bound a subset of possible E-box elements defined by the consensus sequence AACAGATGGT. Inspection of the selected E-box motifs suggests that TAL1/E47 heterodimers will bind with high affinity to the calculated consensus as well as to related E-box elements with modest deviations from the consensus sequence, particularly if those deviations involve flanking residues of the E-box motif. EMSAs confirm that TAL1/E2A heterodimers bind with high affinity to the consensus E-box sequence (Fig. 2 and 5) and to some related E-box elements (data not shown).

At least two forms of TAL1 can be detected in leukemic cells: a full-length gene product (TAL1α, residues 1 to 331) and a truncated polypeptide (TAL1β, residues 176 to 331) (11). Our data suggest that both species have the potential to associate with each of the four known class A proteins to form at least eight distinct bHLH complexes. Each complex evaluated by the CASTing procedure (TAL1α/E47, TAL1β/E47, TAL1β/E2-2, and TAL1β/HEB) selected a subset of E-box elements that can be represented by the same consensus sequence (AACAGATGGT). Therefore, the sequence preferences for DNA recognition by the different TAL1 heterodimers appear to be indistinguishable. The present results do not establish, however, whether these complexes bind DNA with the same or with differing affinities. Clearly, it remains to be determined whether the various TAL1 heterodimers serve distinct and/or redundant functions during normal development. It is also uncertain whether the leukemic properties of TAL1 are influenced differentially by dimerization with distinct class A proteins.

Class A bHLH proteins have been detected in cell lines derived from T-ALL patients. For example, Jurkat cells express E2A but not E2-2 polypeptides (2). The presence of potential dimerization partners for TAL1 in T-ALL cells is also indicated by the fact that leukemic cell extracts confer upon TAL1 the ability to select E-box elements during the CASTing procedure (Fig. 7). Significantly, the consensus sequence calculated from these elements is identical to the preferred recognition site of TAL1 heterodimers involving the known class A bHLH proteins (AACAGATGGT). These results suggest that the leukemic function of TAL1 is mediated by association with known class A proteins or with other highly related bHLH polypeptides. Indeed, the EMSAs indicate that most of the DNA-binding activity of TAL1 in Jurkat and RPMI8402 T-ALL cells is mediated by heteromeric bHLH complexes involving E2A polypeptides (e.g., Fig. 8). Nevertheless, a minor fraction of the TAL1 complexes is resistant to disruption by E2A-specific antisera and therefore may represent TAL1 heterodimers involving other class A polypeptides such as HEB.

Half-site recognition of DNA by TAL1 and the class A bHLH proteins. It has been proposed that each subunit of a bHLH dimer interacts primarily with one half of the E-box recognition sequence (7). This notion, supported by recent crystallographic studies of protein-DNA complexes (13), implies that the E-box motif can be viewed as a combination of two half-sites. The CASTing procedure, and other binding-site selection techniques, have been used to identify optimal sequences for DNA recognition by heterodimers consisting of a myogenic bHLH factor (MyoD1 or myogenin) and a class A protein (E12 or E47) (7, 45). The preferred sequence bound by these heterodimers (e.g., AACACCTGTT) was interpreted as a combination of the half-site recognized by the class A protein (AACAC) and the half-site recognized by the myogenic factor (CTGTT). We were initially unable to assign half-sites to the preferred recognition sequence for TAL1 heterodimers (AACAGATGGT) since it does not contain the proposed E2A half-site (AACAC). Therefore, the sequence recognition properties of E2A homodimers were further evaluated by CASTing experiments with E47 polypeptides (Fig. 7). As noted previously by Blackwell and Weintraub (7), the preferred binding sequence derived from these experiments (AACAGGTGTT) is asymmetric in the central residues of the E-box motif. However, the preferred sequence at the flanking residues of the motif is symmetric and therefore allowed the calculation of two possible E47 half-sites (AACAG and AACAC). One of these is present in the preferred recognition sequence for

TAL1 heterodimers (AACAGATGGT). Hence, the latter can be viewed as a half-site for class A polypeptides (AACAG) combined with a half-site for TAL1 polypeptides (ATGGT). Notably, the TAL1 half-site is distinct from that recognized by myogenic class B proteins (CTGTT) (7, 45).

E2A polypeptides serve as components of several bHLH complexes, including E2A homodimers, myogenic heterodimers, and TAL1 heterodimers (16, 23, 26, 34, 35, 38, 44). However, it appears that the DNA recognition properties of an E2A polypeptide are influenced significantly by its dimerization partner. Thus, whereas the E2A half-site recognized by myogenic heterodimers is AACAC, the preferred E2A half-site of TAL1 heterodimers is AACAG. Indeed, the asymmetric nature of the optimal binding sequence for E2A homodimers (AACAGGTGTT) can be accounted for if one polypeptide recognizes the AACAG half-site while the other preferentially binds the AACAC half-site (the complement of which is GTGTT). Clearly, the preferred half-site recognized by an E2A polypeptide (e.g., AACAG or AACAC) is determined, at least in part, by its bHLH partner.

Potential targets for transcriptional regulation by TAL1. Although malignant activation of *TAL1* is associated with T-cell leukemia, *TAL1* expression also occurs in a variety of normal settings, including hematopoietic cells within the erythroid, megakaryocytic, and mastocytic lineages (33), as well as in restricted populations of endothelial cells (25). Like other bHLH proteins, TAL1 may serve in vivo as a transcription factor. If so, it will be necessary to identify the subordinate genes subject to regulation by TAL1 in order to fully appreciate its role during either normal or malignant development. Identification of the preferred sequences for DNA recognition by TAL1 heterodimers represents a first step in that direction. Significantly, the optimal binding sequences for myogenic heterodimers are well represented within the transcriptional control regions of muscle-specific genes, including those that have been implicated as regulatory targets of myogenic bHLH proteins (7, 45). Likewise, subordinate genes regulated by TAL1 may harbor sequences within their transcriptional regulatory elements that are related to the preferred TAL1 recognition site. Indeed, a preliminary survey of the nucleotide data base revealed possible TAL1-binding sites associated with a number of candidate target genes. For example, potential TAL1-binding sequences are found in the U3 regions of most mammalian type C retroviruses (AACAGATGGT) (19, 37), the serum response element of c-*fos* (CGCAGATGTC) (32), the 3′ untranslated region of the endothelial ELAM-1 gene (GACAGATGTT) (22), erythroid promoters of the carbonic anhydrase II (TTCATATGTT), band 3 (TCCAGATGTG), and GATA-1 (ATCATATGTA) genes (12, 28, 48), in hypersensitivity site 2 of the β-globin locus control region (CCCAGATGTT) (40), and near the 3′ enhancer of the human $^{A}\gamma$-globin gene (AACAGATGTT) (8). It remains to be seen whether these or other related sequences represent bona fide sites for in vivo recognition by TAL1 heterodimers.

ACKNOWLEDGMENTS

We thank Tom Kadesch and Cornelis Murre for generously providing plasmid clones. We are also grateful to Susan Campbell for preparing the manuscript.

This work was supported by grant CA46593 to R.B. from the National Cancer Institute. J.-S.H. and R.E.K. were supported by a grant from Hoechst AG. R.B. is a member of the Simmons Arthritis Research Center and the recipient of faculty research award FRA-421 from the American Cancer Society.

REFERENCES

1. **Aplan, P. D., D. P. Lombardi, A. M. Ginsberg, J. Cossman, V. L. Bertness, and I. R. Kirsch.** 1990. Disruption of the human SCL locus by "illegitimate" V-(D)-J recombinase activity. Science **250:**1426–1429.
2. **Bain, G., S. Gruenwald, and C. Murre.** 1993. E2A and E2-2 are subunits of B-cell-specific E2-box DNA-binding proteins. Mol. Cell. Biol. **13:**3522–3529.
3. **Bash, R. O., W. M. Crist, J. J. Shuster, M. P. Link, M. Amylon, J. Pullen, A. J. Carroll, G. R. Buchanan, R. G. Smith, and R. Baer.** 1993. Clinical features and outcome of T-cell acute lymphoblastic leukemia in childhood with respect to alterations at the *TAL1* locus: a Pediatric Oncology Group Study. Blood **81:**2110–2117.
4. **Begley, C. G., P. D. Aplan, S. M. Denning, B. F. Haynes, T. A. Waldmann, and I. R. Kirsch.** 1989. The gene SCL is expressed during early hematopoiesis and encodes a differentiation-related DNA-binding motif. Proc. Natl. Acad. Sci. USA **86:**10128–10132.
5. **Bernard, O., P. Guglielmi, P. Jonveaux, D. Cherif, S. Gisselbrecht, M. Mauchauffe, R. Berger, C.-J. Larsen, and D. Mathieu-Mahul.** 1990. Two distinct mechanisms for the SCL gene activation in the t(1;14) translocation of T-cell leukemias. Genes Chrom. Cancer **1:**194–208.
6. **Bernard, O., N. Lecointe, P. Jonveaux, M. Souyri, M. Mauchauffe, R. Berger, C. J. Larsen, and D. Mathieu-Mahul.** 1991. Two site-specific deletions and t(1;14) translocation restricted to human T-cell acute leukemias disrupt the 5′ part of the *tal-1* gene. Oncogene **6:**1477–1488.
7. **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.
8. **Bodine, D. M., and T. J. Ley.** 1987. An enhancer element lies 3′ to the human $^{A}\gamma$ globin gene. EMBO J. **6:**2997–3004.
9. **Brown, L., J.-T. Cheng, Q. Chen, M. J. Siciliano, W. Crist, G. Buchanan, and R. Baer.** 1990. Site-specific recombination of the *tal*-1 gene is a common occurrence in human T cell leukemia. EMBO J. **9:**3343–3351.
10. **Chen, Q., J.-T. Cheng, L.-H. Tsai, N. Schneider, G. Buchanan, A. Carroll, W. Crist, B. Ozanne, M. J. Siciliano, and R. Baer.** 1990. The *tal* gene undergoes chromosome translocation in T cell leukemia and potentially encodes a helix-loop-helix protein. EMBO J. **9:**415–424.
11. **Cheng, J.-T., H.-L. Hsu, L.-Y. Hwang, and R. Baer.** 1993. Products of the *TAL1* oncogene: basic helix-loop-helix proteins phosphorylated at serine residues. Oncogene **8:**677–683.
12. **Disela, C., C. Glineur, T. Bugge, J. Sap, G. Stengl, J. Dodgson, H. Stunnenberg, H. Beug, and M. Zenke.** 1991. v-*erb*A overexpression is required to extinguish c-*erb*A function in erythroid cell differentiation and regulation of the *erb*A target gene CAII. Genes Dev. **5:**2033–2047.
13. **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.
14. **Field, J., J.-I. Nikawa, D. Broek, B. MacDonald, L. Rodgers, I. A. Wilson, R. A. Lerner, and M. Wigler.** 1988. Purification of a *RAS*-responsive adenylyl cyclase complex from *Saccharomyces cerevisiae* by use of an epitope addition method. Mol. Cell. Biol. **8:**2159–2165.
15. **Finger, L. R., J. Kagan, G. Christopher, J. Kurtzberg, M. S. Hershfield, P. C. Nowell, and C. M. Croce.** 1989. Involvement of the *TCL5* gene on human chromosome 1 in T-cell leukemia and melanoma. Proc. Natl. Acad. Sci. USA **86:**5039–5043.
16. **Funk, W. D., M. Ouellette, and W. E. Wright.** 1991. Molecular biology of myogenic regulatory factors. Mol. Biol. Med. **8:**185–195.
17. **Funk, W. D., D. T. Pak, R. H. Karas, W. E. Wright, and J. W. Shay.** 1992. A transcriptionally active DNA-binding site for human p53 protein complexes. Mol. Cell. Biol. **12:**2866–2871.
18. **Funk, W. D., and W. E. Wright.** 1992. 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. Proc. Natl. Acad. Sci. USA **89**:9484–9488.

19. **Golemis, E. A., N. A. Speck, and N. Hopkins.** 1990. Alignment of U3 region sequences of mammalian type C virus: identification of highly conserved motifs and implications for enhancer design. J. Virol. **64**:534–542.
20. **Guan, K., and J. E. Dixon.** 1991. Eukaryotic proteins expressed in *Escherichia coli*: an improved thrombin cleavage and purification procedure of fusion proteins with glutathione *S*-transferase. Anal. Biochem. **192**:262–267.
21. **Henthorn, P., M. Kiledjian, and T. Kadesch.** 1990. Two distinct transcription factors that bind the immunoglobulin enhancer μE5/κE2 motif. Science **247**:467–470.
22. **Hession, C., L. Osborn, D. Goff, G. Chi-Rosso, C. Vassallo, M. Pasek, C. Pittack, R. Tizard, S. Goelz, K. McCarthy, S. Hopple, and R. Lobb.** 1990. Endothelial leukocyte adhesion molecule 1: direct expression cloning and functional interactions. Proc. Natl. Acad. Sci. USA **87**:1673–1677.
23. **Hsu, H.-L., J.-T. Cheng, Q. Chen, and R. Baer.** 1991. Enhancer-binding activity of the *tal-1* oncoprotein in association with the E47/E12 helix-loop-helix proteins. Mol. Cell. Biol. **11**:3037–3042.
24. **Hu, J.-S., E. N. Olson, and R. E. Kingston.** 1992. HEB, a helix-loop-helix protein related to E2A and ITF2 that can modulate the DNA-binding ability of myogenic regulatory factors. Mol. Cell. Biol. **12**:1031–1042.
25. **Hwang, L.-Y., M. Siegelman, L. Davis, N. Oppenheimer-Marks, and R. Baer.** 1993. Expression of the *TAL1* proto-oncogene in cultured endothelial cells and blood vessels of the spleen. Oncogene **8**:3043–3046.
26. **Kadesch, T.** 1993. Consequences of heteromeric interactions among helix-loop-helix proteins. Cell Growth Differ. **4**:49–55.
27. **Kamps, M. P., C. Murre, X.-H. Sun, and D. Baltimore.** 1990. A new homeobox gene contributes the DNA binding domain of the t(1;19) translocation protein in Pre-B ALL. Cell **60**:547–555.
28. **Kopito, R. R., M. A. Andersson, and H. F. Lodish.** 1987. Multiple tissue-specific sites of transcriptional initiation of the mouse anion antiport gene in erythroid and renal cells. Proc. Natl. Acad. Sci. USA **84**:7149–7153.
29. **Lassar, A. B., R. L. Davis, W. E. Wright, T. Kadesch, C. Murre, A. Voronova, D. Baltimore, and H. Weintraub.** 1991. Functional activity of myogenic HLH proteins requires hetero-oligomerization with E12/E47-like proteins *in vivo*. Cell **66**:305–315.
30. **Lee, K. A. W., A. Bindereif, and M. R. Green.** 1988. A small-scale procedure for preparation of nuclear extracts that support efficient transcription and pre-mRNA splicing. Genet. Anal. Tech. Appl. **5**:22–31.
31. **Mellentin, J. D., S. D. Smith, and M. L. Cleary.** 1989. *lyl*-1, a novel gene altered by chromosomal translocation in T cell leukemia, codes for a protein with a helix-loop-helix DNA binding motif. Cell **58**:77–83.
32. **Metz, R., and E. Ziff.** 1991. The helix-loop-helix protein rE12 and the C/EBP-related factor rNFIL-6 bind to neighboring sites within the c-*fos* serum response element. Oncogene **6**:2165–2178.
33. **Mouthon, M.-A., O. Bernard, M.-T. Mitjavila, P.-H. Romeo, W. Vainchenker, and D. Mathieu-Mahul.** 1993. Expression of *tal*-1 and GATA-binding proteins during human hematopoiesis. Blood **81**:647–655.
34. **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.
35. **Murre, C., P. S. McCaw, H. Vaessin, M. Caudy, L. Y. Jan, Y. N. 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.
36. **Nelson, C., L.-P. Shen, A. Meister, E. Fodor, and W. J. Rutter.** 1990. Pan: a transcriptional regulator that binds chymotrypsin, insulin, and AP-4 enhancer motifs. Genes Dev. **4**:1035–1043.
37. **Nielsen, A. L., N. Pallisgaard, F. S. Pedersen, and P. Jørgensen.** 1992. Murine helix-loop-helix transcriptional activator proteins binding to the E-box motif of the Akv murine leukemia virus enhancer identified by cDNA cloning. Mol. Cell. Biol. **12**:3449–3459.
38. **Olson, E. N.** 1990. MyoD family: a paradigm for development? Genes Dev. **4**:1454–1461.
39. **Rabbitts, T. H.** 1991. Translocations, master genes, and differences between the origins of acute and chronic leukemias. Cell **67**:641–644.
40. **Reddy, P. M. S., and C.-K. J. Shen.** 1991. Protein-DNA interactions *in vivo* of an erythroid-specific, human β-globin locus enhancer. Proc. Natl. Acad. Sci. USA **88**:8676–8680.
41. **Sanger, F., A. R. Coulson, B. G. Barrell, A. J. H. Smith, and B. A. Roe.** 1980. Cloning in single-stranded bacteriophage as an aid to rapid DNA sequencing. J. Mol. Biol. **143**:161–178.
42. **Smith, D. B., and K. S. Johnson.** 1988. Single-step purification of polypeptides expressed in *Escherichia coli* as fusions with glutathione *S*-transferase. Gene **67**:31–40.
43. **Sun, X.-H., N. G. Copeland, N. A. Jenkins, and D. Baltimore.** 1991. Id proteins Id1 and Id2 selectively inhibit DNA binding by one class of helix-loop-helix proteins. Mol. Cell. Biol. **11**:5603–5611.
44. **Weintraub, H., R. Davis, S. Tapscott, M. Thayer, M. Krause, R. Benezra, T. K. Blackwell, D. Turner, R. Rupp, S. Hollenberg, Y. Zhuang, and A. Lassar.** 1991. The MyoD gene family: nodal point during specification of the muscle cell lineage. Science **251**:761–766.
45. **Wright, W. E., M. Binder, and W. Funk.** 1991. Cyclic amplification and selection of targets (CASTing) for the myogenin consensus binding site. Mol. Cell. Biol. **11**:4104–4110.
46. **Xia, Y., L. Brown, C. Y.-C. Yang, J. T. Tsan, M. J. Siciliano, R. Espinosa III, M. M. Le Beau, and R. J. Baer.** 1991. *TAL2*, a helix-loop-helix gene activated by the (7;9)(q34;q32) translocation in human T-cell leukemia. Proc. Natl. Acad. Sci. USA **88**:11416–11420.
47. **Zhang, Y., J. Babin, A. L. Feldhaus, H. Singh, P. A. Sharp, and M. Bina.** 1991. HTF4: a new human helix-loop-helix protein. Nucleic Acids Res. **19**:4555.
48. **Zon, L. I., and S. H. Orkin.** 1992. Sequence of the human GATA-1 promoter. Nucleic Acids Res. **20**:1812.