

Coding Sequences of the *tal-1* Gene Are Disrupted by Chromosome Translocation in Human T Cell Leukemia

By Qi Chen, Cary Ying-Chuan Yang, Julia Tsou Tsan, Ying Xia, Abdelsalam H. Ragab,*^{||} Stephen C. Peiper,[§] Andrew Carroll,^{‡||} and Richard Baer

From the Department of Microbiology, University of Texas Southwestern Medical Center, Dallas, Texas 75235; the *Department of Pediatrics, Emory University Medical Center, Atlanta, Georgia 30322; the [‡]Laboratory of Medical Genetics and the [§]Department of Pathology, University of Alabama, Birmingham, Alabama 35294; and the ^{||}Pediatric Oncology Group, St. Louis, Missouri 63108

Summary

The *tal-1* proto-oncogene encodes a helix-loop-helix DNA-binding protein that has been implicated in the formation of T cell acute lymphoblastic leukemia (T-ALL). Patients with T-ALL harbor structural rearrangements of *tal-1* that result from either local DNA deletion or t(1;14)(p34;q11) chromosome translocation. By analyzing t(1;14)(p34;q11) chromosomes from a series of patients, we have now identified a discrete region of *tal-1* wherein most of the translocation breakpoints occur. Moreover, mapping of *tal-1* genomic DNA revealed that coding exons are situated on both sides of the t(1;14)(p34;q11) major breakpoint region. Hence, the translocated allele of *tal-1* is truncated in a manner that reduces its amino acid coding potential.

The *tal-1* gene was identified upon analysis of t(1;14)(p34;q11), a chromosome translocation observed in the malignant cells of patients with T cell acute lymphoblastic leukemia (T-ALL)¹ (1–4). Although it is found in only 3% of these patients, t(1;14)(p34;q11) has been implicated as a causative factor of leukemogenesis based on its occurrence in unrelated patients and its exclusive association with T-ALL (5). The translocation generates a reciprocal exchange of genetic material between chromosomes 1 and 14 with cytogenetic breakage in chromosomal bands 1p34 and 14q11. Consequently, the *tal-1* gene from chromosome 1 is transposed into the TCR- α/δ chain locus at 14q11, whereupon its expression is presumably altered in a manner that promotes the formation of T-ALL.

Two additional lines of evidence support the notion that *tal-1* is a protooncogene that potentially contributes to the development of T-ALL. First, the *tal-1* gene product is a member of the helix-loop-helix (HLH) family of DNA binding proteins (1, 6), several of which have been shown to specifically recognize core sequences of eucaryotic transcriptional enhancers (7–10). Although HLH proteins control diverse aspects of normal cell growth and differentiation, three members of this family (i.e., *c-myc*, *lxl-1*, and *E2A*) have

also been implicated in human acute lymphoblastic leukemia (11–13). Second, despite the infrequency of t(1;14)(p34;q11) translocation, we have recently shown that ~25% of T-ALL patients harbor a site-specific 90-kb deletion (*tal^d*) of the *tal-1* gene that is apparently undetectable by cytogenetic analysis (14). Therefore, structural alteration of *tal-1*, either by t(1;14)(p34;q11) translocation or *tal^d* deletion, represents the most common genetic lesion associated with human T-ALL.

The mechanism by which translocated alleles of *tal-1* are activated in malignancy is not understood. To address this issue, we have analyzed the structures of t(1;14)(p34;q11) chromosome junctions from a series of T-ALL patients. In each of five cases, the translocation breakpoint on chromosome 1 occurred within a discrete 1-kb region of the *tal-1* locus. Moreover, mapping of *tal-1* genomic DNA revealed that coding exons are situated on both sides of the t(1;14)(p34;q11) major breakpoint region (mbr). Hence, the translocated allele of *tal-1* is truncated in a manner that reduces its amino acid coding potential.

Materials and Methods

The Patients. Clinical features of the pediatric T-ALL patients (cases 1 and 2) and karyotypic studies of their leukemic cells have already been described (5). The adult T-ALL patient (case 0) was a 42-yr-old male with marked hyperleukocytosis comprised of leu-

¹ Abbreviations used in this paper: HLH, helix-loop-helix; mbr, major breakpoint region; T-ALL, T cell acute lymphoblastic leukemia.

kemic lymphoblasts. Cytogenetic analysis of his bone marrow lymphoblasts revealed an abnormal karyotype: 46,XY,del(6)(q11q21),t(1;14)(p34;q11). Leukemic cells analyzed in this report were derived from peripheral blood obtained just before treatment. The original designation of the t(1;14) translocation described a breakpoint in the p32 band of chromosome 1 (1); with improved karyotypic resolution of additional cases, we now believe that the breakpoint lies in the proximal portion of p34 rather than in p32 (5).

DNA Analysis and Cloning. DNAs extracted from leukemic specimens were analyzed by Southern hybridization with radiolabeled DNA probes (15). Genomic DNA libraries of BamHI-digested leukemic DNA were constructed in phage vector λ 2001, and re-

striction fragments of recombinant λ DNA were subcloned into plasmid and M13 phage vectors (15). Nucleotide sequence analyses were performed on M13 single-stranded templates by the chain terminator method (16).

Results and Discussion

In a recent cytogenetic study, the t(1;14)(p34;q11) translocation was observed in 5 of 168 children with T-ALL (5). We previously characterized translocation junctions from two of these patients (cases 4 and 5), and found that the chromo-

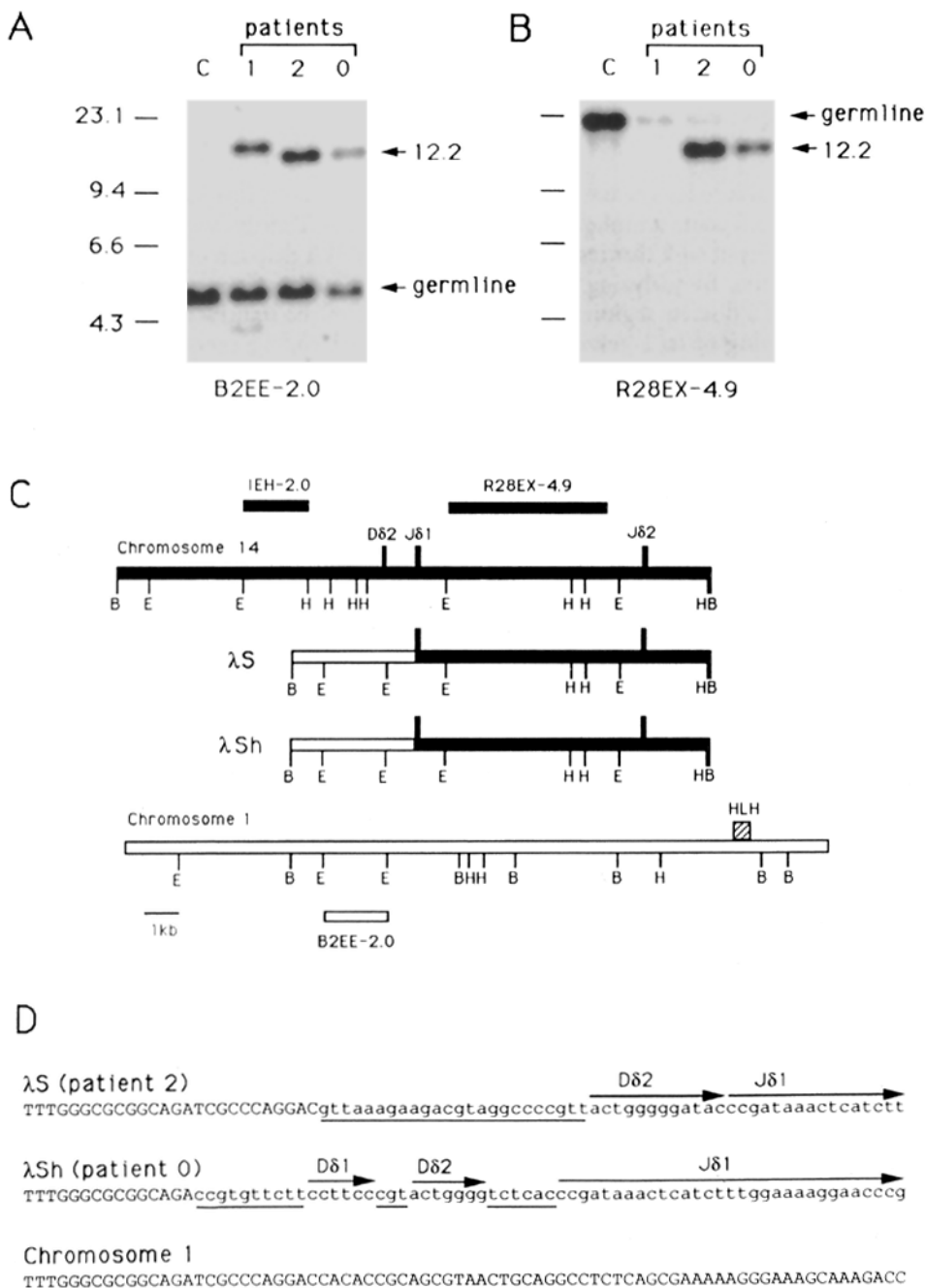


Figure 1. The t(1;14)(p34;q11) junctions of patients 0 and 2. Southern analysis of BamHI-digested DNAs hybridized with probe B2EE-2.0 (A) or probe R28EX-4.9 (B). (Lanes C) Control DNA from a B lymphoblastoid cell line; (lanes 1, 2, and 0) leukemic DNA from patients 1, 2, and 0, respectively. Sizes of HindIII λ DNA fragments are indicated (in kilobases) to the left of each autoradiogram. (C) Recombinant clones λ S and λ Sh were obtained from phage libraries of leukemic DNAs from patients 2 and 0, respectively. The λ S and λ Sh maps are compared with those of the germline *tal-1* locus and the D δ -J δ region of a germline TCR- α/δ chain gene. Closed boxes represent chromosome 14 sequences, and open boxes represent chromosome 1 sequences. Sequences encoding the *tal-1* HLH domain are represented by a hatched box. (D) Comparison of t(1;14)(p34;q11) junction sequences from patients 2 and 0 with the germline sequence of chromosome 1. Uppercase letters represent chromosome 1 sequence, and lowercase letters represent chromosome 14 sequence (excepting nucleotides generated by random N-region insertion, which are underlined). Coding sequences derived from the TCR gene segments D δ 1, D δ 2, and J δ 1 are overlined with arrows.

some 1 breakpoints occurred 1 kb apart at positions ~11 and ~12 kb, respectively, upstream of genomic sequences encoding the *tal-1* HLH domain (see Fig. 1 C) (1). To determine whether this region represents a common site of chromosome translocation, we have analyzed *t(1;14)(p34;q11)* leukemic DNAs from an adult T-ALL patient (case 0) and two additional pediatric patients (cases 1 and 2). As shown in Fig. 1 A, patients 0 and 2 both exhibit a rearranged 12.2-kb BamHI fragment upon Southern hybridization with the *tal-1* DNA probe B2EE-2.0. Moreover, rearranged fragments of identical size were also detected in these patients by hybridization with a DNA probe (R28EX-4.9) from the J δ region of the TCR- α/δ chain gene (Fig. 1 B). Therefore, λ phage libraries of BamHI-digested leukemic DNAs from patients 0 and 2 were constructed, and recombinant clones containing the 12.2-kb fragments were isolated by simultaneous screening with B2EE-2.0 and R28EX-4.9. Since the rearranged 12.2-kb fragments contain sequences from both chromosome 1 (B2EE-2.0) and chromosome 14 (R28EX-4.9), they are likely to encompass the *t(1;14)(p34;q11)* junctions of patients 0 and 2. This was confirmed by restriction mapping and nucleotide sequence analysis of the recombinant clones (λ Sh from patient 0; λ S from patient 2), which revealed that the rearranged fragments are comprised of *tal-1* sequences juxtaposed with the TCR J δ 1 gene segment (Fig. 1, C and D).

As illustrated in Fig. 1 A, the *tal-1* probe B2EE-2.0 de-

fects two rearranged BamHI fragments (14.1 and 4.3 kb) in leukemic DNA from patient 1. To determine whether either fragment represents a *t(1;14)(p34;q11)* junction, a λ phage library of BamHI-digested DNA from patient 1 was constructed and screened with the B2EE-2.0 probe. In this manner, three hybridizing clones were obtained, each of which contained the rearranged 14.1-kb fragment. Restriction mapping and sequence analysis of one such clone (λ A) revealed it to be comprised of B2EE-2.0 sequences (chromosome 1) juxtaposed with sequences from the C δ region of the TCR- α/δ chain gene (chromosome 14) (Fig. 2).

Two abnormal chromosomes, der(1) and der(14), are generated as a consequence of *t(1;14)(p34;q11)* translocation (5). From the known orientation of the TCR- α/δ chain gene on chromosome 14, we can deduce that the *t(1;14)(p34;q11)* junctions of clones λ S, λ Sh, and λ A (Figs. 1 C and 2 A) are derived from the der(1) chromosomes of patients 2, 0, and 1, respectively. A genetic map of *tal-1* is provided in Fig. 3, and the der(1) breakpoints from these patients are illustrated along with those previously determined for patients 4 and 5. The *t(1;14)(p34;q11)* breakage in patient 1 occurs within sequences represented by the B2EE-2.0 probe; hence, Southern analysis with this probe revealed two rearranged fragments in digests of patient 1 leukemic DNA (Fig. 1 A), one of which corresponds to the der(1) junction (i.e., the 14.1-kb BamHI fragment), while the other is likely to represent

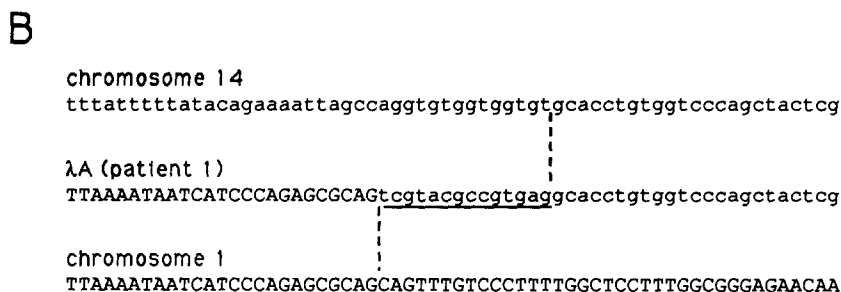
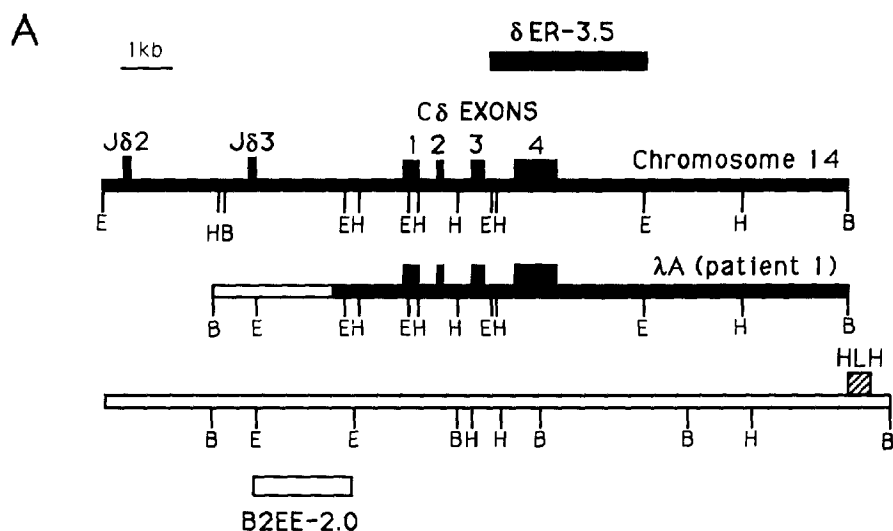


Figure 2. The *t(1;14)(p34;q11)* junction of patient 1. (A) Recombinant clone λ A was obtained from a phage library of patient 1 leukemic DNA. The λ A map is compared with those of the germline *tal-1* locus and the J δ -C δ region of the TCR- α/δ chain gene. Closed boxes represent chromosome 14 DNA, and open boxes represent chromosome 1 DNA. Sequences encoding the *tal-1* HLH domain are represented by a shaded box. (B) Comparison of the *t(1;14)(p34;q11)* junction sequence with germline sequences from chromosomes 1 and 14. Uppercase letters represent chromosome 1 sequence, and lowercase letters represent chromosome 14 sequence (excepting nucleotides generated by random N-region insertion, which are underlined).

the der(14) junction (i.e., the 4.3-kb BamHI fragment). Southern hybridization with the *tal-1* probe SU2PP (Fig. 3) revealed DNA rearrangements corresponding to the der(14) junctions in leukemic DNA from each of the five patients (data not shown). Hence, in these patients, the t(1;14)(p34;q11) rearrangement is conservative with respect to chromosome 1; that is, the continuity of *tal-1* was disrupted, but few, if any, *tal-1* sequences were lost as a consequence of the translocation. In contrast, Southern analyses with probes located upstream of the chromosome 14 breakpoints (e.g., EH-2.0; Fig. 1 C) indicate that substantial portions of the TCR locus were lost from the translocated alleles (data not shown), in accord with previous studies of other chromosome defects involving the TCR- α/δ chain gene (17).

As illustrated in Fig. 3, the five t(1;14)(p34;q11) breakpoints described in our studies cluster within a 1-kb region located ~11 kb upstream of sequences encoding the *tal-1* HLH domain. Recently, t(1;14)(p34;q11) translocations from two additional patients were described at the molecular level (2-4). Notably, t(1;14)(p34;q11) breakage in the T-ALL cell line Kd also occurs within the same region of *tal-1* that is disrupted in patients 0, 1, 2, 4, and 5 (2). In contrast, the t(1;14)(p34;q11) breakpoint of DU.528, a cell line derived from a rare stem cell leukemia, lies within 3'-noncoding sequences (designated SCL or tcl-5) of the *tal-1* transcription unit (3, 4). In sum, six of the seven t(1;14)(p34;q11) translocations characterized

to date feature breakage within what can be described as the major breakpoint region (mbr) of *tal-1*.

Since t(1;14)(p34;q11) is observed in only 3% of T-ALL patients (5), alteration of *tal-1* by chromosome translocation is clearly an uncommon factor in the formation of T-ALL. Nevertheless, we recently detected tumor-specific rearrangement of the *tal-1* gene in ~25% of T-ALL patients, including those without apparent karyotypic lesions of chromosome 1 (14). Surprisingly, the *tal-1* rearrangements observed in different patients are nearly identical; that is, they all result from a precise 90-kb deletion (designated *tal^d*) that arises independently in each patient by site-specific recombination (14). It may be significant that the downstream endpoint of the *tal^d* deletion occurs just 1 kb upstream of the *tal-1* mbr (see Fig. 3). Hence, *tal^d* deletion and t(1;14)(p34;q11) translocation are structurally analogous lesions in that both serve to remove upstream sequences from the body of the *tal-1* gene.

To understand the effect of t(1;14)(p34;q11) translocation on *tal-1* gene expression, it is necessary to localize the mbr with respect to the *tal-1* transcription unit. Although the initiation site(s) of *tal-1* transcription is not known, at least two distinct mRNA species have been identified by cDNA sequence analysis (1, 6). Both transcripts encode the *tal-1* HLH domain but differ from each other in their 5' sequences. To determine the origin of these sequences, we examined *tal-1* genomic DNA by Southern hybridization with oligonucle-

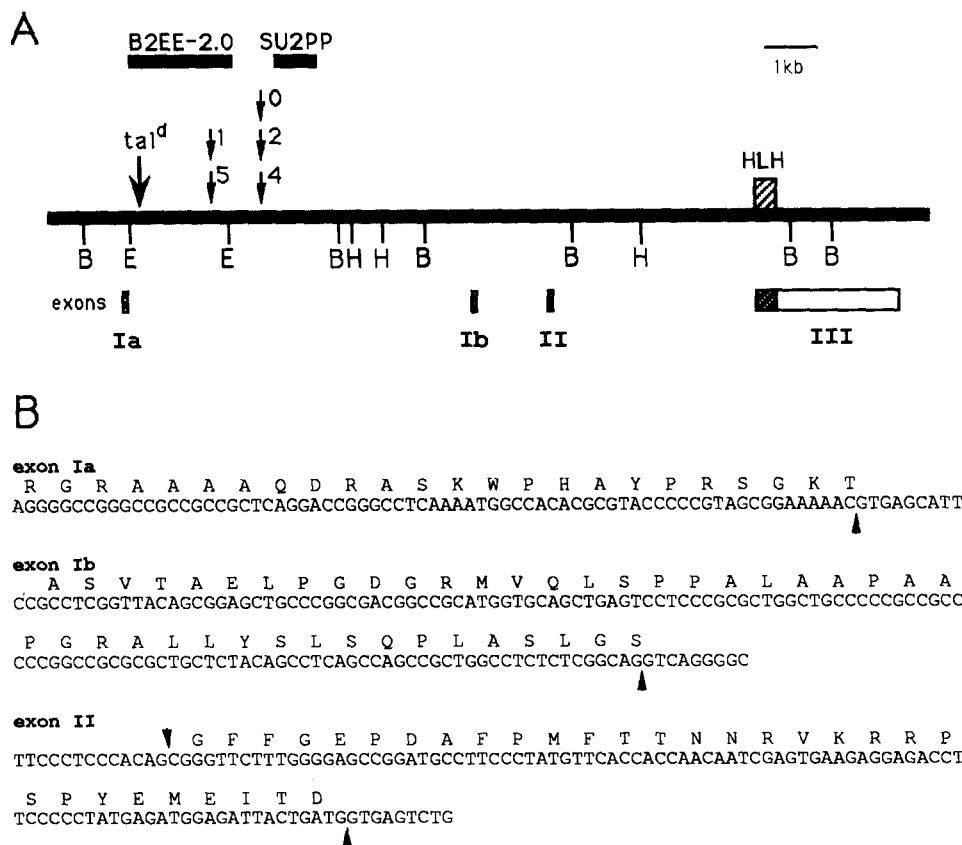


Figure 3. (A) Genetic map of the *tal-1* locus. The small arrows designate the breakpoints of t(1;14)(p34;q11) translocations from patients 0, 1, 2, 4, and 5. The large arrow indicates the downstream endpoint of the 90-kb *tal^d* deletion. The hatched regions of the *tal-1* exons denote coding sequences, and the open regions denote 3' noncoding sequence. Restriction sites: B, BamHI; E, EcoRI; H, HindIII. (B) Genomic DNA sequences encompassing the known upstream exons (Ia, Ib, and II) of *tal-1*. RNA splice sites are indicated with arrowheads.

otide probes representing *tal-1* cDNA sequences. These studies revealed that transcribed sequences of *tal-1* are derived from at least four exons (provisionally designated Ia, Ib, II, and III), the genomic positions and nucleotide sequences of which are illustrated in Fig. 3. The distinct *tal-1* transcripts are apparently derived as a consequence of alternative RNA splicing, since one species (designated type A mRNA) includes exons Ia, II, and III, while the other (type B mRNA) contains exons Ib, II, and III. The effect of t(1;14)(p34;q11) translocation on the expression of type B mRNA transcripts cannot be evaluated until its transcription start site is defined. Nevertheless, since the t(1;14)(p34;q11) mbr lies downstream of exon Ia, it is clear that translocated alleles of *tal-1* are usually truncated so as to preclude normal expression of type A transcripts (Fig. 3).

Sequences encoding the *tal-1* HLH domain are located within exon III at a position ~11 kb downstream of the t(1;14)(p34;q11) mbr (Fig. 3). The cDNA sequences determined for type A and type B transcripts do not extend sufficiently 5' to include the initiator codons of protein synthesis (1, 6). Hence, known sequences within the upstream exons (Ia, Ib, and II) also have amino acid coding potential. The existence of coding sequences within exon Ia is especially significant because this exon is removed from the *tal-1* locus as a consequence of t(1;14)(p34;q11) translocation. Moreover, the 90-kb deletion engendered by *tal*^h recombination also results in the removal of exon Ia (14). Consequently, the *tal-1* gene alterations associated with T-ALL, whether they arise by t(1;14)(p34;q11) translocation or *tal*^h recombination, are structurally analogous lesions that eliminate normal expression of the type A mRNA protein product.

The linkage between *tal-1* gene truncation and T-ALL is unclear. Nevertheless, studies of other oncogenes that encode presumptive DNA-binding proteins provide some intriguing precedents. For example, most oncogenically activated myb proteins are truncated in a manner that removes an NH₂-terminal site for phosphorylation by casein kinase II; it has been proposed that the loss of this site promotes neoplasia by releasing Myb from normal physiological controls (18). The studies of *c-myc* and *lyl-1* may be especially relevant to the mechanism of *tal-1* activation since these genes also encode HLH proteins implicated in human acute lymphoblastic leukemia. Protein translation from normal *c-myc* mRNA begins at either of two alternative sites in the same reading frame: a leucine codon (CUG) from exon 1 or a methionine codon (AUG) from exon 2. Hence, the normal *c-myc* gene encodes a 67-kd polypeptide initiated at leucine (*c-Myc-1*), as well as a 64-kd species initiated at methionine (*c-Myc-2*). A nearly universal feature of oncogenically activated *c-myc* genes is the mutation or removal of exon 1 sequences necessary for production of *c-Myc-1* (19). Loss of exon 1 coding sequences has also been implicated in the malignant activation of *lyl-1* (12). Although the biochemical functions of sequences encoded by the upstream exons of *c-myc*, *lyl-1*, and *tal-1* are not known, disruption of these sequences is clearly a common occurrence in human leukemogenesis. This is especially prominent in T-ALL, where >25% of patients bear altered alleles of *tal-1* in which exon Ia is removed as a result of either t(1;14)(p34;q11) translocation or *tal*^h recombination.

We thank Cindy Baselski for preparing the manuscript. We are also grateful to Luigi Bertoli and William Crist for their advice.

This work was supported by the National Cancer Institute (CA-46593, CA-47975, CA-20549, and CA-30969). R. Baer is a member of the Simmons Arthritis Research Center and a recipient of a Junior Faculty Research Award from the American Cancer Society. A. H. Ragab is an American Cancer Society Professor of Clinical Oncology. Q. Chen and Y. Xia are students in the Microbiology Program at University of Texas Southwestern Graduate School.

Address correspondence to Richard Baer, Department of Microbiology, University of Texas Southwestern Medical Center, 5323 Harry Hines Blvd., Dallas, TX 75235.

Received for publication 9 July 1990.

References

1. 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 (Eur. Mol. Biol. Organ.) J.* 9:415.
2. 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, Chromosomes & Cancer.* 1:194.
3. 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.
4. Begley, C.G., P.D. Aplan, M.P. Davey, K. Nakahara, K. Tchorz, J. Kurtzberg, M.S. Hershfield, B.F. Haynes, D.I. Cohen, T.A.

- Waldmann, and I.R. Kirsch. 1989. Chromosomal translocation in a human leukemia stem-cell line disrupts the T-cell antigen receptor δ -chain diversity region and results in a previously unreported fusion transcript. *Proc. Natl. Acad. Sci. USA.* 86:2031.
5. Carroll, A., W. Crist, M.P. Link, M.D. Amylon, D.J. Pullen, A.H. Ragab, G.R. Buchanan, R.S. Wimmer, and T.J. Vietti. 1990. The t(1;14)(p34;q11) is non-random and restricted to T-cell acute lymphoblastic leukemia. *Blood.* In press.
 6. 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.
 7. 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.
 8. 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.
 9. 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.
 10. Henthorn, P., M. Kiledjian, and T. Kadesch. 1990. Two distinct transcription factors that bind the immunoglobulin enhancer μ E5/ κ E2 motif. *Science (Wash. DC).* 247:467.
 11. Champlin, R., and R.P. Gale. 1989. Acute lymphoblastic leukemia: recent advances in biology and therapy. *Blood.* 73:2051.
 12. 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.
 13. Mellentin, J.D., C. Murre, T.A. Donlon, P.S. McCaw, S.D. Smith, A.J. Carroll, M.E. McDonald, D. Baltimore, and M.L. Cleary. 1989. The gene for enhancer binding proteins E12/E47 lies at the t(1;19) breakpoint in acute leukemias. *Science (Wash. DC).* 246:379.
 14. 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 (Eur. Mol. Biol. Organ.) J.* In press.
 15. Sambrook, J., E.F. Fritsch, and T. Maniatis, Eds. 1989. *Molecular Cloning: A Laboratory Manual.* Second Edition. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
 16. 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.
 17. Baer, R., K.-C. Chen, S.D. Smith, and T.H. Rabbitts. 1985. Fusion of an immunoglobulin variable gene and a T cell receptor constant gene in the chromosome 14 inversion associated with T cell tumors. *Cell.* 43:705.
 18. Lüscher, B., E. Christenson, D.W. Litchfield, E.G. Krebs, and R.N. Eisenman. 1990. Myb DNA binding inhibited by phosphorylation at a site deleted during oncogenic activation. *Nature (Lond.).* 344:517.
 19. Hann, S.R., M.W. King, D.L. Bentley, C.W. Anderson, and R.N. Eisenman. 1988. A non-AUG translational initiation in *c-myc* exon 1 generates an N-terminally distinct protein whose synthesis is disrupted in Burkitt's lymphomas. *Cell.* 52:185.