Proc. Natl. Acad. Sci. USA
Vol. 92, pp. 12160–12164, December 1995
Medical Sciences

Domains with transcriptional regulatory activity within the ALL1 and AF4 proteins involved in acute leukemia

(chromosome translocations/11q23 abnormalities/activator/repressor)

R. PRASAD*†, T. YANO*†, C. SORIO*, T. NAKAMURA*, R. RALLAPALLI*, Y. GU*, D. LESHKOWITZ‡, C. M. CROCE*, AND E. CANAANI‡

*Jefferson Cancer Institute and Department of Microbiology and Immunology, Thomas Jefferson University, Philadelphia, PA 19107; and ‡Department of Molecular Cell Biology, Weizmann Institute of Science, Rehovot 76100, Israel

Communicated by Raymond L. White, University of Utah, Salt Lake City, UT, August 18, 1995

ABSTRACT **The *ALL1* gene, located at chromosome band 11q23, is involved in acute leukemia through a series of chromosome translocations and fusion to a variety of genes, most frequently to *AF4* and *AF9*. The fused genes encode chimeric proteins. Because the *Drosophila* homologue of *ALL1*, trithorax, is a positive regulator of homeotic genes and acts at the level of transcription, it is conceivable that alterations in *ALL1* transcriptional activity may underlie its action in malignant transformation. To begin studying this, we examined the ALL1, AF4, AF9, and AF17 proteins for the presence of potential transcriptional regulatory domains. This was done by fusing regions of the proteins to the yeast GAL4 DNA binding domain and assaying their effect on transcription of a reporter gene. A domain of 55 residues positioned at amino acids 2829–2883 of ALL1 was identified as a very strong activator. Further analysis of this domain by *in vitro* mutagenesis pointed to a core of hydrophobic and acidic residues as critical for the activity. An ALL1 domain that repressed transcription of the reporter gene coincided with the sequence homologous to a segment of DNA methyltransferase. An AF4 polypeptide containing residues 480–560 showed strong activation potential. The C-terminal segment of AF9 spanning amino acids 478–568 transactivated transcription of the reporter gene in HeLa but not in NIH 3T3 cells. These results suggest that ALL1, AF4, and probably AF9 interact with the transcriptional machinery of the cell.**

Chromosome 11q23 abnormalities are associated with 5–10% of human acute leukemias, in particular in children under the age of 1 year, and in secondary leukemias (1). These leukemias have a lymphoid, myeloid, or lymphoid–myeloid phenotype and uniformly carry a poor prognosis. More than 20 different 11q23 translocations have been identified by now (2, 3). In nearly all of these aberrations, the *ALL1* gene (also called *HRX, MLL*, or *HTRX*) is involved (4–7). *ALL1* is thought to be the human homologue of the *Drosophila* trithorax gene, which acts as a positive regulator of the homeotic gene complexes Antennapedia and Bithorax.

11q23 chromosome translocations cleave *ALL1* within the breakpoint cluster region and fuse it to a partner gene, resulting in production of fused RNAs and presumably of chimeric proteins (4, 5). The chimeric proteins containing the N terminus of ALL1 are thought to be the critical products of the translocations (for review, see ref. 8). Presently, 10 partner genes have been cloned (8–13).

The mechanism(s) by which the chimeric ALL1 proteins act to trigger neoplasm is not known. Two models were proposed: the first (5) emphasized gain of function and suggested that the altered protein (specifically ALL1/ENL) is a hybrid transcription factor in which the DNA binding domain of ALL1 localizes the hypothetical transactivation domain of ENL to regulatory sites of target genes. The second (14) suggested that the cleavage of ALL1 protein results in loss of its function and that the partner polypeptide augments this loss. Resolution between the two models would necessitate developing biological and biochemical assays to measure transcriptional regulatory activity of the normal ALL1 protein, of the normal partner proteins, and of the chimeric species. As a first step in this direction, we looked for domains with transcriptional activation or repression potential within ALL1, within the AF4 and AF9 partner proteins (presumed to be nuclear proteins), and within the AF17 partner polypeptide, which contains a dimerization motif. A similar approach to investigate the ALL1 protein was recently reported (15).

MATERIALS AND METHODS

Plasmids. Effector plasmid vectors were either RSV-GAL4 (16) or pGALM (17). The former contains a GAL4 DNA binding domain (amino acids 1–147) flanked by the Rous sarcoma virus (RSV) promoter and simian virus 40 (SV40) splice and polyadenylylation signals. pGALM is similar to the first effector with the exception that the promoter is the SV40 early promoter. Reporter plasmids were either GAL4-TKCAT in which five GAL4 binding sites are placed adjacent to the thymidine kinase (TK) TATA box, CCAAT element, Sp1 sites, and the chloramphenicol acetyltransferase (CAT) reporter gene (18), or GAL4-E1bCAT plasmid containing five GAL4 DNA binding sites, the E1bTATA sequence, and the CAT gene (19). The β-galactosidase plasmid CH110 was obtained from Pharmacia. All constructs were examined for expression of the GAL4 chimeric proteins by transfection into COS-7 cells and subsequent Western blot analysis of equal amounts of proteins utilizing rabbit anti-GAL4 antiserum (Upstate Biotechnology).

Site-Directed Mutagenesis. Mutagenesis was performed by two-stage PCR amplification (20). The products were cloned into the pGALM effector vector and sequenced.

Transfections. Exponentially growing Chinese hamster ovary cells (CHO-DUKX) in 100-mm dishes were transfected by the DEAE-dextran method with 2 μg each of the pGALM derivatives, GAL4-E1bCAT, and the CH110 plasmid. Cells were collected 48 hr after dimethyl sulfoxide shock and chloroquine treatment, disrupted by three cycles of freezing and thawing, and assayed for CAT and β-galactosidase activities. Results of CAT assays were normalized according to the β-galactosidase values. HeLa and NIH 3T3 cells were transfected by the calcium phosphate method. Effector plasmids were the derivatives of RSV-GAL4 or pGALM (1 μg). Re-

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

Abbreviations: RSV, Rous sarcoma virus; SV40, simian virus 40; TK, thymidine kinase; CAT, chloramphenicol acetyltransferase; STRD, strong transrepression domain.

†R.P. and T.Y. contributed equally to this work.

porter plasmids consisted of GAL4-TKCAT or GAL4-E1bCAT (1 μg). Five micrograms of reference β-galactosidase plasmid was also included. CAT activity was determined by TLC or by extraction of the product into organic solvents. Transfections with all constructs were repeated at least three times.

RESULTS

cDNAs coding for segments of the ALL-1 protein (Fig. 1) were ligated in-frame to the GAL4 DNA binding domain within the effector plasmid vectors RSV-GAL4 (F1, F2, F9, F10) and pGALM (F3–F8). The RSV-GAL4 and pGALM constructs were cotransfected into HeLa cells together with the GAL4-TKCAT reporter. Aliquots of the transfected cells were assayed for CAT activity; the results were normalized for β-galactosidase activity and compared to the results obtained from transfections with insert-less effector vectors. One ALL1 domain, spanning amino acids 1032–1395, repressed expression of the reporter gene. A second domain encompassing amino acids 2772–3579 activated transcription of the CAT gene (Fig. 1). These results were reproducible in HeLa and NIH 3T3 cells. The transrepression activity could be demonstrated only with the GAL4-TKCAT reporter because of a relatively high basal activity. When using the GAL4-E1bCAT reporter, which has a low basal activity, we could demonstrate, in transfections into CHO cells, 300- to 500-fold activation by fragment 9 driven by the SV40 promoter (see below). The levels of activation and repression exhibited by the two ALL1 fragments were at least as high as those obtained with VP16 (21) and YY1 (18), respectively (data not shown). Activation and repression were dependent on the presence of GAL4 target sites within the reporters (data not shown). To ensure similar levels of expression of ALL1/GAL4 proteins, the corresponding constructs were transfected into COS-7 cells and assayed by Western blot analysis with a polyclonal antibody against the GAL4 DNA binding domain. All constructs showed comparable levels of protein expression; some of this analysis is shown in Fig. 1*B*.

A similar analysis, yielding results close to ours, was recently reported (15). However, here we were able to demonstrate transactivation potency only in the region spanning amino acids 2772–3579, and not in the region flanked by amino acids 2339–2772, reported as a potent activator in HeLa cells (15). The latter region was found in our study to be inactive in HeLa, NIH 3T3, or CHO cells, regardless of the reporters (GAL4-E1bCAT or GAL4-TKCAT). Moreover, constructs containing *ALL1* segments F6–F8, which together span amino acids 2332–2807, did not show any activity with either of the reporters.

To delineate the ALL1 transactivation domain, the region containing amino acids 2604–3579 was divided by PCR amplification into a series of overlapping segments (Fig. 2*A*), which were cloned into the pGALM vector and assayed by cotransfection with GAL4-E1bCAT into CHO cells. The P2 region, which was found positive, was further dissected into a variety of fragments and similarly assayed. This analysis identified a minimal transactivation domain as an ALL1 segment containing amino acids 2829–2883. This segment showed 300- to 500-fold activation, while the overlapping segments of amino acids 2808–2864, 2829–2864, 2844–2883, 2829–2853, and 2853–2883 indicated 220-, 85-, 125-, 3-, and 1-fold activation, respectively (data not shown). The absence of any activity in the last two constructs suggested that residues adjacent to amino acid 2853 are critical. To identify these residues, we mutated single or groups of amino acids and assayed the products for CAT activity (Fig. 2*B*). Substitution of the two aspartic acids at positions 2848 and 2851 by asparagines (M5) resulted in loss of ≈90% of the activity. Substitution of the

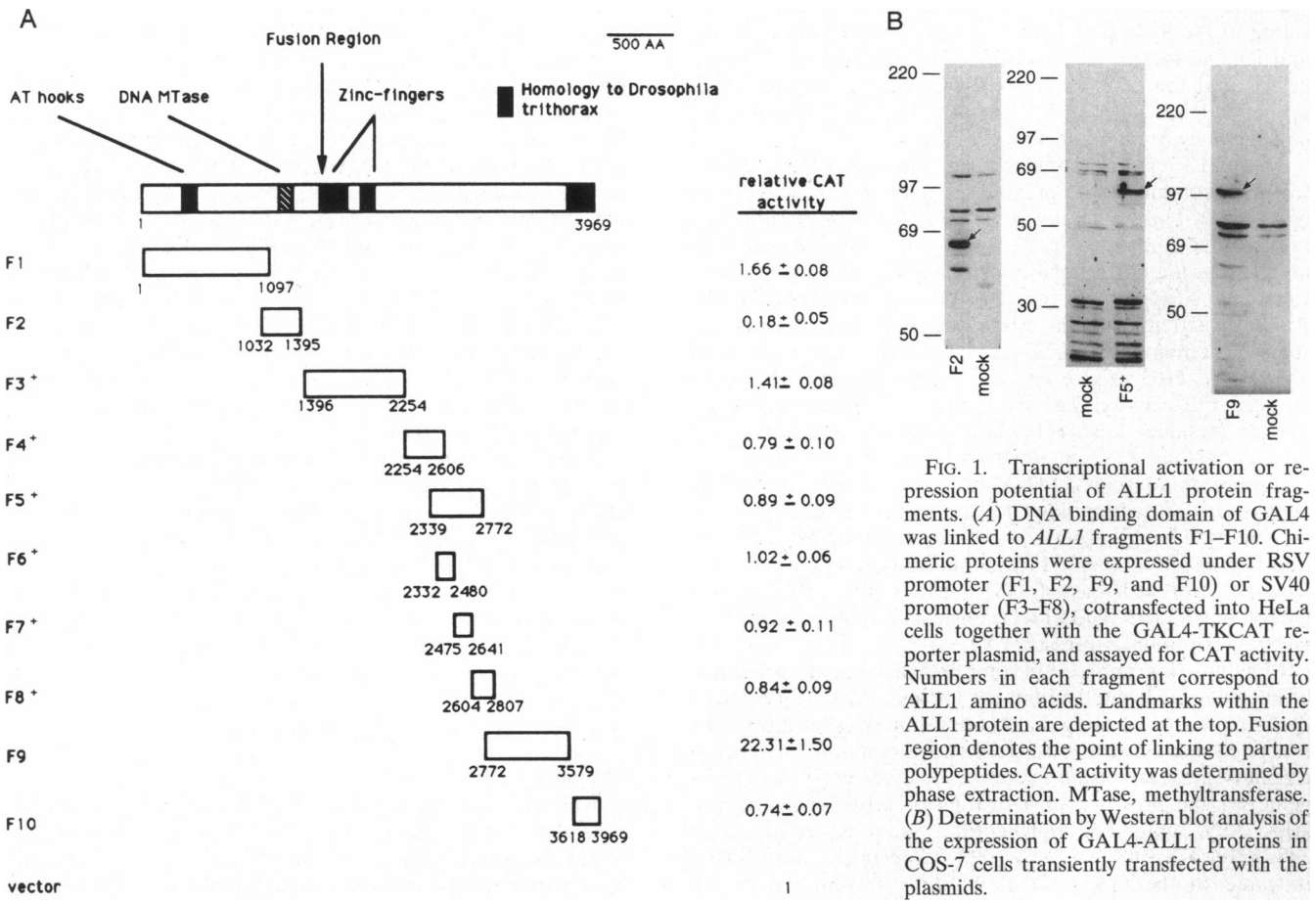
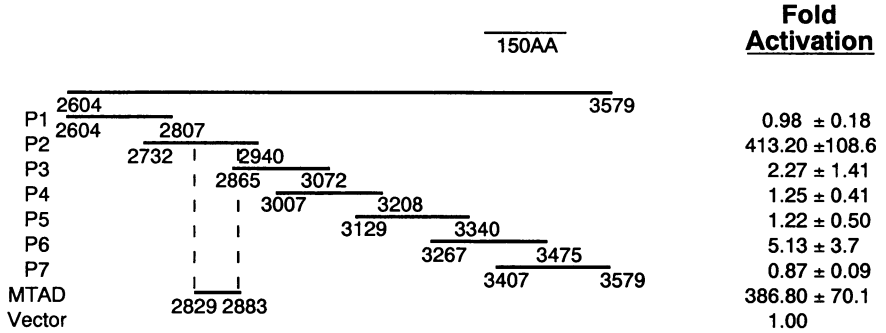


FIG. 1. Transcriptional activation or repression potential of ALL1 protein fragments. (*A*) DNA binding domain of GAL4 was linked to *ALL1* fragments F1–F10. Chimeric proteins were expressed under RSV promoter (F1, F2, F9, and F10) or SV40 promoter (F3–F8), cotransfected into HeLa cells together with the GAL4-TKCAT reporter plasmid, and assayed for CAT activity. Numbers in each fragment correspond to ALL1 amino acids. Landmarks within the ALL1 protein are depicted at the top. Fusion region denotes the point of linking to partner polypeptides. CAT activity was determined by phase extraction. MTase, methyltransferase. (*B*) Determination by Western blot analysis of the expression of GAL4-ALL1 proteins in COS-7 cells transiently transfected with the plasmids.



B

mutants	amino acid sequence	relative activity(%)
MTAD(wt)	LKSDSDNNNSDDCGNILPSDIMDFVLKNTPSMQALGESPESSSSELLNLGEGLGL	100
M 1(DD/NN1)	···N·N·····	75-100
M 2(DD/NN2)	·····NN·····	75-100
M 3(IL/AA)	·····AA·····	10-15
M 4(S/A)	·····A·····	100-140
M 5(DD/NN3)	·····N·N·····	7-15
M 6(IM/AA)	·····AA·····	5-7
M 7(F/A)	·····A·····	30-50
M 8(L/A)	·····A·····	40-60
M 9(L/R)	·····R·····	25-35
M10(FVL/AAA)	·····AAA·····	5-7
M11(HP/5A)	·····AA·AAA·····	1-2
M12(T/D)	·····D·····	75-100
M13(S/H)	·····H·····	100-120
M14(E/Q)	·····Q··Q·····	80-100
M15(4S/4A)	·····AAAA·····	100-160
M16(3L/3A)	·····AA·A·····	160-200

FIG. 2. Localization of the minimal transactivation domain (MTAD) in the ALL1 protein (*A*), and CAT activity of MTAD variants prepared by site-directed mutagenesis (*B*). Constructs were prepared in the pGALM effector vector. Reporter plasmid was GAL4-E1bCAT, and transfections were made into CHO cells. CAT activity was determined by phase extraction.

hydrophobic residues isoleucine and leucine positioned at residues 2844 and 2845 (M3) or isoleucine and methionine at residue 2849 and 2850 (M6) by alanines abolished 85–95% of the activity. Similar loss of activity was observed after substitution of the hydrophobic amino acids phenylalanine, valine, and leucine located at residues 2852–2854 (M10). This suggested that the core of the ALL1 activation domain is the sequence IL(PS)DIMDFVL composed of two aspartic acids and seven hydrophobic residues.

The ALL1 region encompassing amino acids 1032–1395 was found to contain transrepression activity, in agreement with others (15). This segment was dissected by PCR amplification into smaller domains (Fig. 3). Significant activity was distributed across the entire region including sequences originating from exon 4 (e4), exon 5 (e5), and exons 6 and 7 (e6, e7). The strongest activity though, was localized to the e4 construct spanning amino acids 1032–1251. Cleaving this domain at amino acid 1163 to give left and right segments (e4L, e4R) reduced the activity. The e4 sequence includes the ALL1 domain (residues 1153–1219) homologous to a region within the mammalian DNA methyltransferase (8). The fragment spanning amino acids 1147–1240 and containing the DNA methyltransferase homology domain indeed showed the highest activity, 12-fold repression, and was termed STRD (strong transrepression domain). Fragments containing only the cysteine-rich (mtC) or basic (mtB) portions (Fig. 3*B*) were less potent than STRD. The STRD construct was more active than the transcriptional repressor YY1.

AF4 and AF9 represent partner proteins that are presumed or known to be located within the nucleus. AF17 represents a partner protein carrying a potential dimerization motif. We examined the segments of these proteins contained within the chromosome translocation products that are presumed oncogenic for the presence of activating or repressing domains. AF4, AF9, and AF17 polypeptides spanning amino acids 348–1210, 478–568, and 552–1093, respectively, were linked in-frame to the GAL4 DNA binding domain within the RSV-GAL4 effector vector. The plasmids encoding the three partner polypeptides were found to be expressed at comparable levels when transfected into COS-7 cells and assayed by Western blotting (data not shown). The three plasmids were cotransfected with the GAL4-TKCAT reporter into NIH 3T3 cells (Fig. 4 *Left*). To control for GAL4-mediated activities, the same plasmids were also cotransfected with the TKCAT plasmid, lacking the GAL4 responsive element (Fig. 4 *Right*). Effector plasmids encoding the GAL4-VP16 activator, GAL4-YY1 repressor, or the DNA binding domain of GAL4 alone served as controls. AF4 showed strong activity similar to that of VP16; both activities were mediated by GAL4 DNA binding sites (Fig. 4). Weak activity of AF9 in NIH 3T3 cells was not mediated through GAL4 DNA binding sites (Fig. 4); however, when assayed in HeLa cells, AF9 showed 36.6- and 2.9-fold activation of the GAL4-TKCAT and TKCAT reporters, respectively (data not shown). AF17 was found to be negative in both NIH 3T3 cells (Fig. 4) and HeLa cells (data not shown). Neither AF4 nor AF9 transactivated the GAL4-E1bCAT reporter (data not shown).

To map more precisely the activation domain within the AF4 partner polypeptide, we divided it first into four fragments af4-1, af4-2, af4-3, and af4-4 (Fig. 5), inserted them into the pGALM vector, and cotransfected with the GAL4-TKCAT reporter into NIH 3T3 cells. af4-1 and af4-4 showed strong and weak activity, respectively. Fragment af4-1 was further dissected and found to carry a domain at residues 347–460 with a weak activity and a region at 480–560 that was as potent as the entire partner polypeptide (Fig. 5).

DISCUSSION

The ALL1 activation domain identified here was shown to activate two different reporters in HeLa, NIH 3T3, and CHO cells; it was at least as strong as the VP16 transactivator. The minimal transactivating domain contained 55 amino acids. Systematic substitution of amino acids in this domain indicated

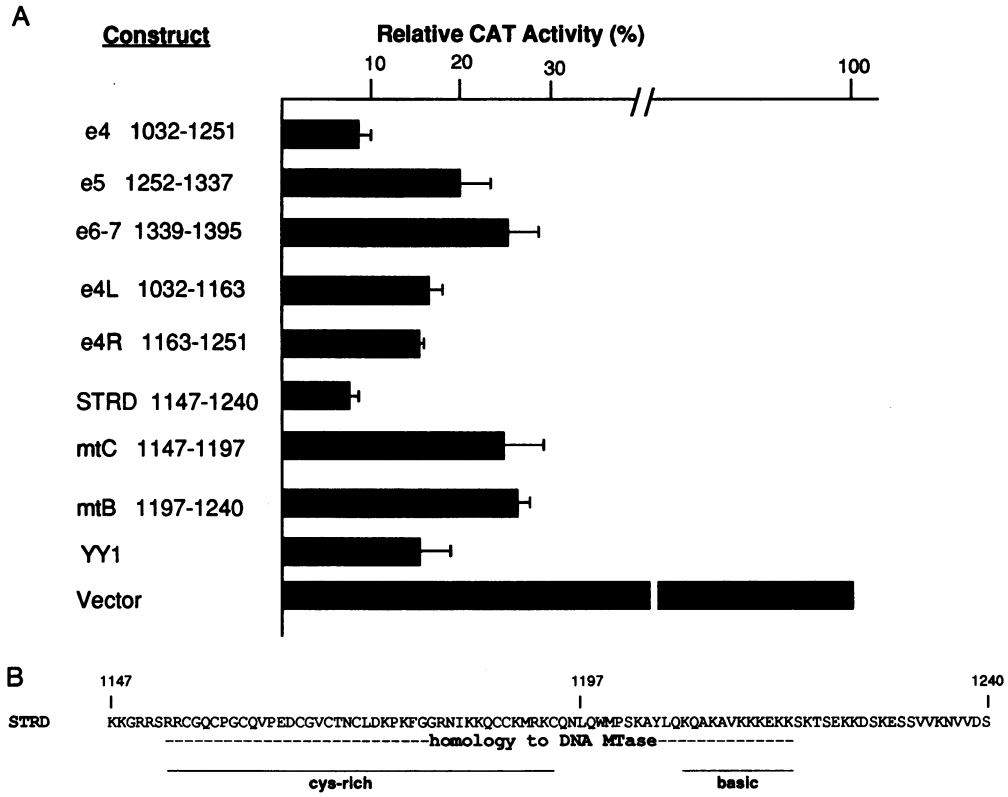


FIG. 3. Fine mapping of the transcriptional repression activity in ALL1 protein (*A*), and sequence of the STRD (*B*). Constructs were made in the RSV-GAL4 effector vector, and the reporter was the GAL4-TKCAT. Transfections were made into NIH 3T3 cells and activities were determined by TLC.

that a core of 9 amino acids, IL(PS)DIMDFVL, is essential, and alterations within the acidic and hydrophobic residues abolish most of the activity. Activation domains in some other proteins share features of this core; in VP16 several aspartic acid residues interspersed with hydrophobic amino acids contribute significantly to the activation function, and substitution of phenylalanine or leucines in that region abolish its function (22). Similar results were observed in analysis of the activation domain of Epstein–Barr virus (23). Finally, bulky hydrophobic amino acids adjacent to glutamines were identified in the Sp1 domains that interact with $TAF_{II}110$; alterations of the hydrophobic residues disrupted the interaction (24). Conceivably,

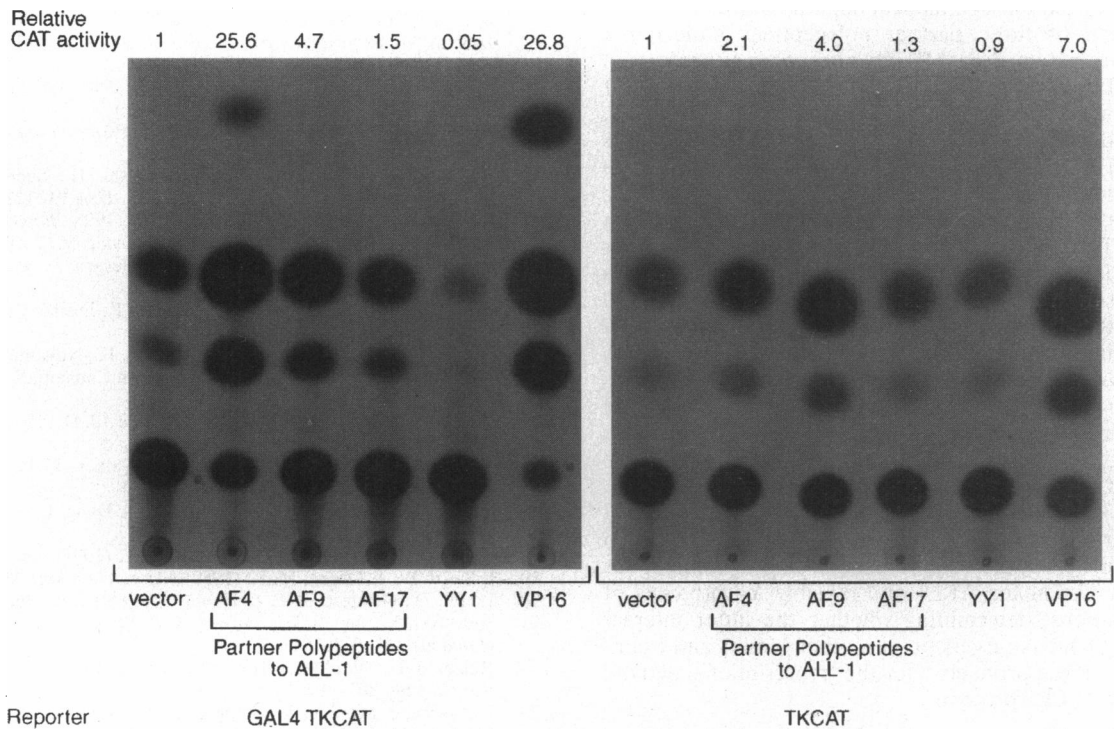


FIG. 4. CAT activity of the AF4, AF9, and AF17 partner polypeptides. Partner polypeptides (corresponding to the segments of the three proteins fused to the ALL1 protein as a consequence of 11q23 translocation) were linked to the GAL4 DNA binding domain residues 1–147 within the RSV-GAL4 vector. Plasmids were cotransfected into NIH 3T3 cells together with the GAL4-TKCAT (*Left*) or TKCAT (no GAL4 binding sites) (*Right*) reporters. Activities were determined by TLC. Vectors containing the YY1 and VP16 proteins served as controls.

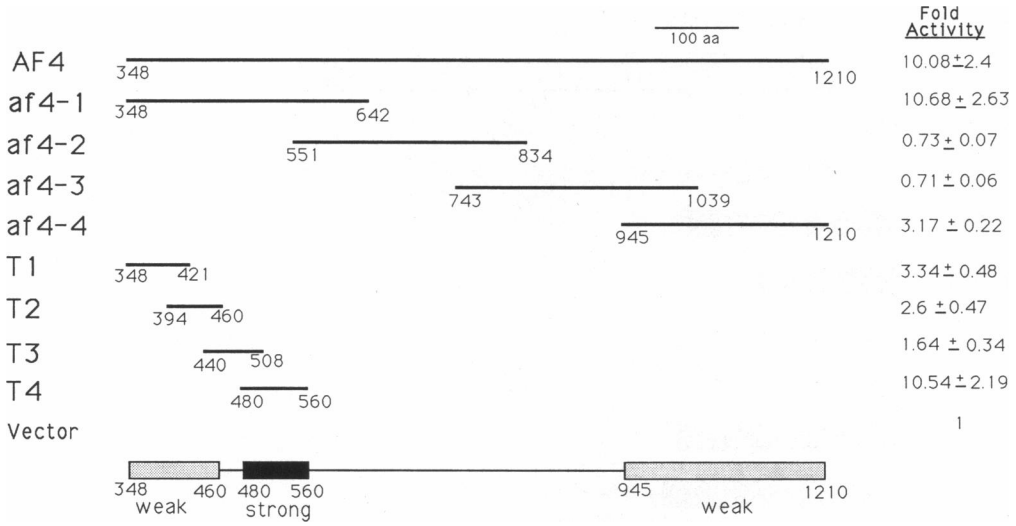


FIG. 5. Fine mapping of the transactivating domain in the AF4 partner polypeptide. Effector and reporter vectors were pGALM and GAL4-TKCAT, respectively. Transfections were made into NIH 3T3 cells and activities were determined by TLC.

the core of the ALL1 activation domain forms a structure recognized by the TATA-binding protein-associated factors (TAFs) or by a basal transcription factor. The activation domain will be absent from the products of 11q23 translocations, presumed to be oncogenic.

The second element in ALL1, which we identified by CAT assays, has a transcriptional repression activity. This activity distributes over a region of ≈350 amino acids, but the strongest activity maps to a domain of 93 amino acids, which spans the DNA methyltransferase homologous sequence. The physiological relevance of the repression activity described here is not known. The domain identified would be present in the oncogenic products of 11q23 translocations and included in the segment(s) of ALL1 protein, which is duplicated in some patients in whom *ALL1* is rearranged through tandem duplication (8).

Examination of three partner polypeptides indicated a strong activation domain in AF4 when assays were performed in either NIH 3T3 or HeLa cells. The C-terminal domain of AF9 was highly active in HeLa cells but not in NIH 3T3 cells. AF17 showed no activity. Unlike most activators, the domains present in AF4 and AF9 are reporter dependent; they activate the TK promoter, which includes TATA box, CAAT element, and Sp1 sites, but they fail to activate the E1b promoter, which has only a TATA box. This suggests that the two partner polypeptides interact with a transcription element, which requires other responsive sequences in addition to a TATA box. It is also possible that the polypeptides interact with a component required only for certain promoters. Similar promoter selectivity was recently shown for an activation domain located at the C terminus of the ENL partner polypeptide (25). The C termini of AF9 and ENL share high sequence homology and are presumably the portions of the two proteins essential for oncogenicity of the chimeric ALL1/AF9 and ALL1/ENL proteins.

This work demonstrates that elements of ALL1, AF4, and AF9 affect transcription in a model heterologous system. To study this further would require establishing a homologous transcription system for ALL1 and possibly for products of 11q23 aberrations, determining whether the latter interact with the ALL1 DNA targets or with other targets and examining whether these products alter the transcriptional activity of the normal ALL1 protein.

We are indebted to Drs. M. Walker, C. V. Dang, A. T. Hoang, and S. Reddy for their support and help and to Drs. Y. Shi and T. Shenk for plasmids. These studies were supported by National Cancer Institute Grants CA50507 and CA39860, by a grant from the Tobacco Research Council, and by grants from the Israel Academy of Science and the U.S.–Israel Binational Fund.

1. Mitelman, F. (1994) *Catalogue of Chromosome Aberrations in Cancer.* (Wiley–Liss, New York), 5th Ed.
2. Thirman, M. J., Gill, H. J., Burnett, R. C., Mbangkollo, D., McCabe, N. R., Kobayashi, H., Ziemin-Van der Poel, S., Kaneko, Y., Morgan, R., Sandberg, A. A., Chaganti, R. S. K., Larson, R. A., LeBeau, M. M., Diaz, M. O. & Rowley, J. D. (1993) *N. Engl. J. Med.* **329,** 909–914.
3. Hunger, S. P., Tkachuk, D. C., Amylon, M. D., Link, M. P., Carroll, A. J., Welborn, J. L., Willman, C. L. & Cleary, M. L. (1993) *Blood* **81,** 3197–3203.
4. Gu, Y., Nakamura, T., Alder, H., Prasad, R., Canaani, O., Cimino, G., Croce, C. & Canaani, E. (1992) *Cell* **71,** 701–708.
5. Tkachuk, D. C., Kohler, S. & Cleary, M. L. (1992) *Cell* **71,** 691–700.
6. Ziemin-van der Poel, S., McCabe, N. R., Gill, H. J., Espinosa, R., III, Patel, Y. D., Harden, A. M., LeBeau, M. M., Smith, S. B., Rowley, J. D. & Diaz, M. O. (1991) *Proc. Natl. Acad. Sci. USA* **88,** 10735–10739.
7. Djabali, M., Selleri, L., Parry, P., Bower, M., Young, B. D. & Evans, G. A. (1992) *Nat. Genet.* **2,** 113–118.
8. Canaani, E., Nowell, P. C. & Croce, C. M. (1995) *Adv. Cancer Res.* **66,** 213–234.
9. Parry, P., Wei, Y. & Evan, G. (1994) *Genes Chromosomes Cancer* **11,** 79–84.
10. Thirman, M. J., Levitan, D. A., Kobayashi, H., Simon, M. C. & Rowley, J. D. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 12110–12114.
11. Tse, W., Zhu, W., Chen, H. S. & Cohen, A. (1995) *Blood* **85,** 650–656.
12. Chaplin, T., Ayton, P., Bernard, O. A., Saha, V., Della Valle, V., Hillion, J., Gregorini, A., Lillington, D., Berger, R. & Young, B. D. (1995) *Blood* **85,** 1435–1441.
13. Mitani, K., Kanda, Y., Ogawa, S., Tanaka, T., Inazawa, J., Yazaki, Y. & Hirai, H. (1995) *Blood* **85,** 2017–2024.
14. Prasad, R., Leshkowitz, D., Gu, Y., Alder, H., Nakamura, T., Saito, H., Huebner, K., Berger, R., Croce, C. M. & Canaani, E. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 8107–8111.
15. Zeleznik-Le, N. J., Harden, A. M. & Rowley, J. D. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 10610–10614.
16. Aronheim, A., Shiran, R., Rosen, A. & Walker, M. D. (1993) *Proc. Natl. Acad. Sci. USA* **90,** 8063–8067.
17. Kato, G. J., Barrett, J., Villa-Garcia, M. & Dang, C. V. (1990) *Mol. Cell. Biol.* **10,** 5914–5920.
18. Shi, Y., Seto, E., Chang, L.-S. & Shenk, T. (1991) *Cell* **67,** 377–388.
19. Lillie, J. W. & Green, M. R. (1989) *Nature (London)* **338,** 39–44.
20. Landt, D., Grunert, H.-P. & Hahn, U. (1990) *Gene* **96,** 125–128.
21. Sadowski, I., Ma, J., Triezenberg, S. & Ptashne, M. (1988) *Nature (London)* **335,** 563–564.
22. Reiger, J. L., Shen, F. & Triezenberg, S. J. (1993) *Proc. Natl. Acad. Sci. USA* **90,** 883–887.
23. Hardwick, J. M., Tse, L., Applegern, N., Nicholas, J. & Velinona, M. A. (1992) *Mol. Cell. Biol.* **66,** 5500–5508.
24. Gill, G., Pascal, E., Tseng, Z. H. & Tjian, R. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 192–196.
25. Rubnitz, J. E., Morissey, J., Savage, P. A. & Cleary, M. L. (1994) *Blood* **84,** 1747–1752.