Proc. Natl. Acad. Sci. USA
Vol. 88, pp. 4094–4098, May 1991
Biochemistry

Characterization of the α_4 integrin gene promoter

(DNA sequencing/RNA mapping/DNA transfection)

GLENN D. ROSEN*, THOMAS M. BIRKENMEIER*, AND DOUGLAS C. DEAN*†

*Department of Internal Medicine, Respiratory and Critical Care Division, and †Department of Cell Biology and Physiology, Washington University School of Medicine, St. Louis, MO 63110

Communicated by David M. Kipnis, February 11, 1991

ABSTRACT A cDNA for the α_4 chain of the $\alpha_4\beta_1$ integrin was described previously [Takada, Y., Elices, M. J., Crouse, C. & Hemler, M. E. (1989) *EMBO J.* 8, 1361–1368]. Primer extension analysis indicated that α_4 mRNA extended well beyond the 5′ end of this cDNA. To clone this 5′ sequence, a primer extension cDNA library was constructed, and a cDNA extending an additional 660 base pairs was isolated. This cDNA hybridized to multiple mRNAs in both T and B lymphocytes, but no α_4 mRNA was found in different tissues or in adherent cell lines. A single α_4 gene was detected in a genomic Southern blot when hybridization was done at high stringency; however, additional bands were observed at lower stringency, indicating the presence of α_4-related genes. Some of the different mRNAs that hybridize to the α_4 cDNA may then be the products of these related genes. Analysis of the α_4 genomic sequence revealed a large first exon of 958 base pairs. Interestingly, translation of α_4 initiates from the second ATG in this exon (nucleotide +744). The first ATG (nucleotide +21) is followed by a termination codon 21 amino acids downstream. Such upstream ATG codons have been implicated in translational control of protooncogenes. One major transcriptional start site was identified by using S1 nuclease and primer extension mapping. Consensus sequences for DNA regulatory elements were found upstream of the gene and in exon 1 and intron 1. The α_4 gene 5′ flanking region acted as a promoter in transfection assays. Detailed characterization of the promoter should provide insight into molecular events regulating expression and tissue specificity of α_4.

The β_1 subgroup of the integrin family of adhesion proteins contains at least six cell surface receptors for extracellular matrix proteins (1–4). Each receptor is heterodimeric, containing a common β_1 subunit and a unique α subunit, which dictates ligand specificity.

$\alpha_4\beta_1$ is unique within the subgroup because of its tissue distribution—it is found on lymphoid and myeloid cells but not on adherent cell types—and because of its functional diversity (4). Like other members of the β_1 subgroup, $\alpha_4\beta_1$ mediates attachment of cells to extracellular matrix. Fibronectin (FN) is the extracellular matrix component that serves as a ligand for $\alpha_4\beta_1$ (5). $\alpha_4\beta_1$ binds to a site near the carboxyl terminus of FN that is alternatively spliced in the mRNA (6), suggesting that lymphocyte adhesion to FN can be controlled by the splicing pattern of FN mRNA. The $\alpha_4\beta_1$ binding site is distinct from the Arg-Gly-Asp-Ser (RGDS) sequence that binds $\alpha_5\beta_1$, the well-characterized FN receptor found on the surface of most adherent cell types (7, 8).

$\alpha_4\beta_1$ is also involved in cell–cell interactions. It acts as a receptor on T cells for the ligand, vascular cell adhesion molecule 1 (VCAM-1), which is expressed on the surface of endothelial cells as a result of their exposure to cytokines (9, 10). This interaction is responsible, at least in part, for recruitment of T cells to areas of vascular inflammation. VCAM-1 and FN bind separate sites on $\alpha_4\beta_1$ (10). FN also accumulates at areas of injury (11) and inflammation (12), suggesting that it may act in conjunction with VCAM-1 to recruit lymphocytes to areas of tissue damage.

α_4 is also involved in regulation of the immune response. Antibodies to α_4 inhibit cytolytic T-cell-mediated destruction of target cells, suggesting that α_4 is required for T-cell recognition of target cells (13). The interaction between T-helper and T-suppressor cells is also blocked by an antibody to α_4 (14). In these studies, it was proposed that interaction of $\alpha_4\beta_1$ with a ligand on suppressor cells inhibits suppressor cell proliferation.

In addition, α_4 is also involved in homing of lymphocytes to Peyer's patch high endothelial venules (15, 16). In these studies, α_4 was found complexed with another β subunit, β_p, and it appears that the $\alpha_4\beta_p$ complex is the homing receptor.

Here, we show that the α_4 cDNA‡ hybridizes to multiple mRNAs. Several different genes are detected on low-stringency hybridization of a Southern blot, suggesting that some of these different mRNAs may be products of α_4-related genes. Interestingly, translation of α_4 initiates at the second ATG in exon 1. The first ATG is followed by a termination codon (21 amino acids downstream)—such upstream ATG codons are common in protooncogenes and are found in most members of the *src* family where they are implicated in translational control (17, 18). The α_4 gene promoter is also examined in transfection assays where its activity is compared to that of other promoters.

MATERIALS AND METHODS

Cell Culture and DNA Transfections. Jurkat, Molt-4, and Raji cell lines were maintained in Rosewell Park Memorial Institute medium (RPMI) (1640) in 10% fetal calf serum. Jurkat cells were transfected by electroporation with a BTX Transfector 300 (BTX, San Diego). Approximately 1×10^7 cells were suspended in 0.1 ml of RPMI medium containing 0.5× HeBS (0.5× HeBS is 10 mM Hepes/70 mM NaCl/2.5 mM KCL/0.35 mM Na_2HPO_4/3.0 mM dextrose), salmon sperm DNA at 1.25 mg/ml, and 30 μg of plasmid DNA and subjected to electroporation at 200 V and 950 μF. After 36 hr, chloramphenicol acetyltransferase (CAT) activity was determined (19).

RNA Extraction, Analysis, and Mapping. Poly(A)-RNA isolation, Northern blotting, primer extension, and S1 nuclease mapping have been described (19–21). Northern blots were hybridized at 42°C in 50% (vol/vol) formamide, 4× SSPE (1× SSPE is 0.18 M NaCl/10 mM NaH_2PO_4/1 mM EDTA, pH 7.4), 10× Denhardt's solution (1× Denhardt's solution is 0.02% polyvinylpyrrolidone/0.02% Ficoll/0.02% bovine serum albumin), 0.1% SDS, and sheared denatured

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: CAT, chloramphenicol acetyltransferase; FN, fibronectin.

‡The sequence reported in this paper has been deposited in the GenBank data base (accession no. M62841).

salmon sperm DNA at 0.4 mg/ml. Blots were washed at 65°C in 0.2× SSPE.

Construction and Screening of an α_4 cDNA Library. A synthetic oligonucleotide primer corresponding to sequences between positions +863 and +901 of the α_4 cDNA (see Fig. 4) was used to make a primer extension cDNA library (22). The library was screened with a synthetic oligonucleotide corresponding to sequences between positions +732 and +767 of the α_4 cDNA (see Fig. 4). The largest cDNA insert was subcloned into the plasmid Bluescript SK (Stratagene), and both strands were sequenced with synthetic oligonucleotide primers using the dideoxynucleotide method (23).

Screening a Genomic Library and DNA Characterization. A human lung fibroblast genomic library in λFix (Stratagene) was screened with the ^{32}P-labeled α_4 cDNA (21). DNA restriction fragments were subcloned in Bluescript SK for restriction enzyme mapping and DNA sequencing.

Southern blot analysis (24) was done by using the ^{32}P-labeled cDNA as a probe. Hybridization was done at 42°C in 4× SSPE/50% formamide/10× Denhardt's solution/0.1% SDS/sheared denatured salmon sperm DNA at 0.4 mg/ml. Blots were washed at 60°C in 0.5× SSPE. Low-stringency hybridization was done under similar conditions, but in 30% formamide, and blots were washed at 55°C in 1× SSPE.

Construction of an α_4–CAT Fusion Gene. α_4 gene promoter sequences were amplified, using PCR (25), from the template plasmid pα4-4.3, which contains a 4.3-kilobase (kb) *Eco*RI/*Sal* I genomic fragment. This fragment contains ≈1 kb of the α_4 gene 5′ flanking sequence. The plasmid was linearized by digestion with *Sal* I, and 5 ng was used as a template for PCR. The 3′ oligonucleotide in the amplification corresponded to α_4 gene sequences between positions +107 and +131 (see Fig. 4). This oligonucleotide was synthesized with a *Hin*dIII site on its 5′ end for cloning purposes. The 5′ oligonucleotide corresponded to sequences between positions −389 and −410 with an *Xho* I on its 5′ end. After amplification, the product was digested with *Xho* I and *Hin*dIII and cloned into the corresponding sites in pSKCAT, which contains the CAT structural gene cloned between the *Pst* I (5′ end) and *Bam*HI (3′ end) sites of Bluescript SK.

RESULTS AND DISCUSSION

An α_4 cDNA that hybridized to two mRNAs in the 5- to 6-kb size range was described previously (13). Primer extension analysis indicated that the α_4 mRNAs extend well beyond the 5′ end of this cDNA (data not shown). To clone this additional 5′ sequence, a primer extension cDNA library was constructed. The longest α_4 clone, selected from this library, was sequenced (this cDNA sequence is contained within the genomic sequence shown later in Fig. 4 between nucleotides +67 and +901). The 3′ end of this cDNA overlapped with 174 base pairs (bp) of the 5′ end of the previously described α_4 cDNA (13), while its 5′ end contained 660 bp of additional sequence.

This cDNA was used to probe a Northern blot of poly(A)-RNA from T-cell lines (Fig. 1*A*). Surprisingly, multiple mRNA species ranging in size from 7.0 to 2.1 kb were observed; however, little hybridization was detected in the 5- to 6-kb region where two α_4 mRNAs were reported previously (13). A similar pattern of mRNAs was observed in the B-cell line Raji, but their abundance was much lower than in the T-cell lines (data not shown). No hybridization was detected in adherent cell lines, fetal lung, kidney, brain, or placenta (data not shown), indicating that the different mRNAs share the same lymphoid specificity.

It was of interest to determine if the mRNAs shared identity in different areas of their sequences (i.e., such would be the case with transcripts from the same gene or very related genes) or if similarity was confined to a single area (i.e., a common domain or sequence shared by otherwise unrelated genes). Two different regions of the α_4 cDNA then were used to probe a Northern blot. The results show that multiple mRNA species are detected with each of the probes,

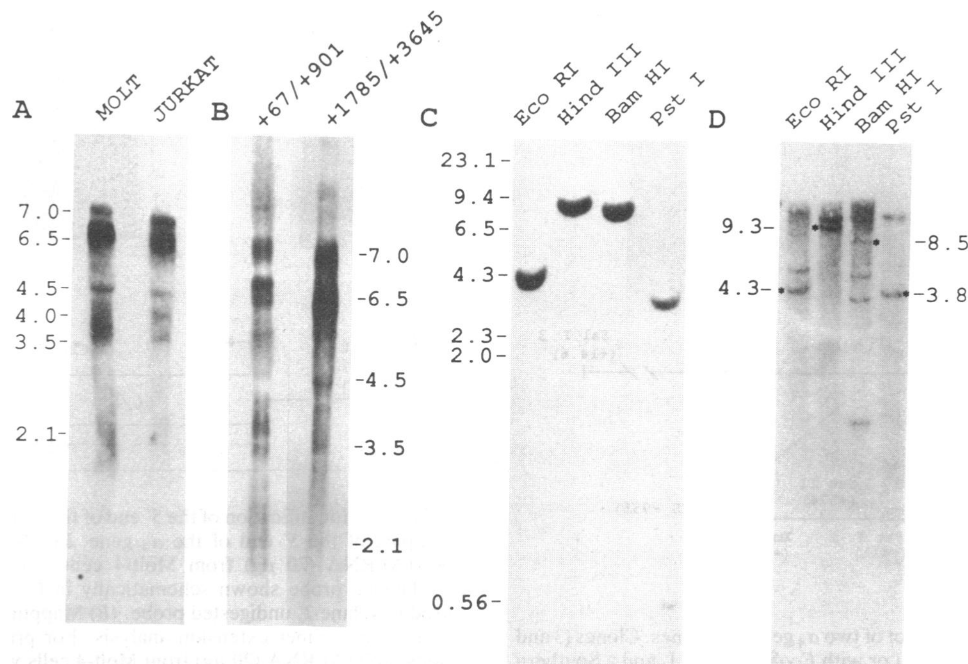

FIG. 1. (*A*) Northern blot showing α_4 mRNAs. A Northern blot containing 10 μg of poly(A)-RNA from Molt-4 and Jurkat cell lines was probed with the α_4 cDNA. (*B*) Hybridization of two different regions of the α_4 cDNA to a Northern blot. Numbers above the lanes indicate the region of α_4 cDNA in the probe. (*C*) Genomic Southern blot of the α_4 gene after digestion with the indicated restriction endonucleases. A Southern blot of human genomic DNA (10 μg per lane) was probed with the α_4 cDNA at high stringency. (*D*) Low-stringency hybridization of the α_4 cDNA to a genomic Southern blot. Asterisks correspond to bands observed at high stringency in *C*. Numbers (in kb) were derived from molecular size standards.

indicating that the different mRNAs are similar over several areas of their sequence (Fig. 1*B*).

The α_4 cDNA detected a single gene in a genomic Southern blot when hybridization was done at high stringency (Fig. 1*C*). However, at lower stringency additional bands were detected (Fig. 1*D*), suggesting the presence of α_4-related genes. Therefore, some of the different mRNA species shown in Fig. 1*A* may be products of these related genes.

The fact that we observed mRNA species that were not seen in a previous study (13) may result from our use of poly(A)-RNA, which facilitates detection of less abundant species. As stated above, we detected little hybridization in the region between 5 and 6 kb, where α_4 mRNAs were identified previously. This could result from discrepancies in determining transcript size or nonspecific hybridization to the 28S ribosomal RNA (which is in this size range) when total RNA was examined.

To determine the genomic structure of the 5′ end and flanking region of the α_4 gene, clones containing the α_4 gene were obtained by screening a genomic library with our α_4 cDNA. A Southern blot of two different clones is shown in Fig. 2*A*. A partial restriction enzyme map of clone 16 is shown in Fig. 2*B*. The 4.3-kb *Eco*RI/*Sal* I fragment, which hybridized to our α_4 cDNA, was subcloned and partially sequenced. The genomic sequence was identical to that of the cDNA up to the point of the first intron, and the sequence identity resumed at the border of exon 2, indicating that the clone contains the α_4 gene.

A combination of S1 nuclease and primer extension mapping was used to define the 5′ end of the gene (Fig. 3). Three closely spaced bands (one major and two minor) mapped to the same location in both primer extension and S1 nuclease assays, suggesting that they each represent transcriptional start sites. The more predominant site was designated +1 (Fig. 3*B*). Identical initiation sites were observed in another T-cell line, the Jurkat cell line (data not shown).

The first ATG in exon 1 occurs at nucleotide +21 (Fig. 4); however, the open reading frame following this ATG extends only 21 amino acids (termination codons are present in each reading frame), indicating that it is not the translational initiation site for α_4. Instead, translation of α_4 initiates at the second ATG (nucleotide +743). Such upstream ATG codons are present in >65% of protooncogenes—they are found in five out of seven members of the *src* gene family (17). As in the α_4 gene, these upstream ATG codons are found within very long 5′ untranslated regions. Removal of upstream ATG codons in the *lck* protooncogene activated translation and was associated with *lck*-mediated oncogenic transformation (18). Therefore, the upstream ATG in the α_4 gene may play a role in translational control.

The entire 743 bp of the α_4 gene 5′ untranslated region is found in exon 1, which is 958 bp long. Consensus sequences for DNA regulatory elements are found in this exon and in the 5′ flanking region—an AP-2 site (+590), a TATA box (−41), PU boxes (−54, −67), a CAAT box (−212), a MyoD binding

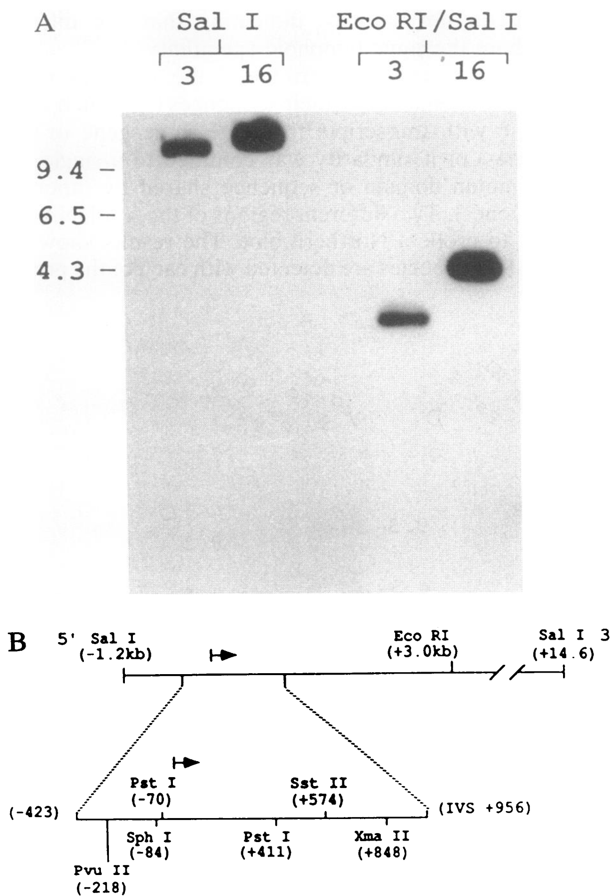

FIG. 2. (*A*) Southern blot of two α_4 genomic clones. Clones (3 and 16) were digested with *Sal* I or with *Eco*RI and *Sal* I, and a Southern blot of the resulting fragments was probed with the α_4 cDNA. Molecular sizes (in kb) are given at left. (*B*) Partial restriction enzyme map of the 5′ end and flanking region of the α_4 gene. The top line shows a map of the 16.4-kb *Sal* I fragment from clone 16, whereas the bottom line shows the portion of this fragment that was sequenced (Fig. 4). An arrow indicates the major transcriptional start site (+1) (Fig. 3). IVS, intervening sequence 1.

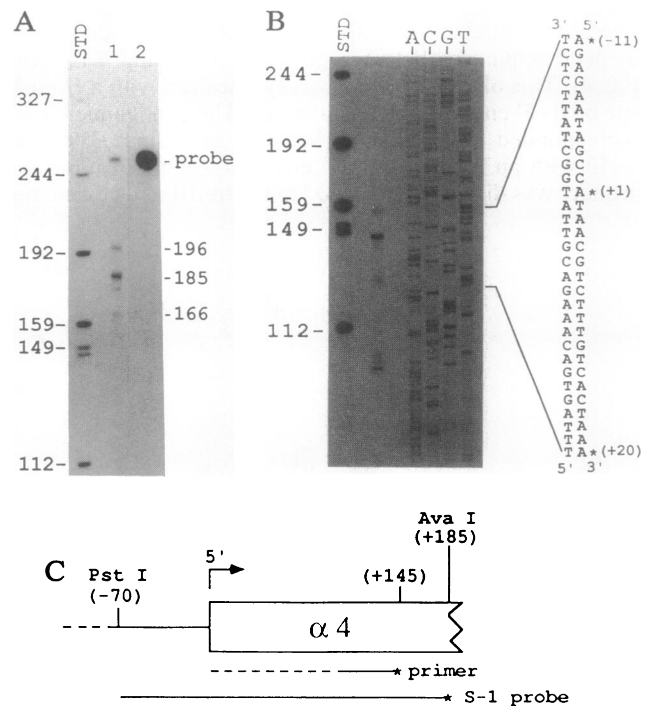

FIG. 3. Identification of the 5′ end of the α_4 gene. (*A*) S1 nuclease mapping of the 5′ end of the α_4 gene. For S1 nuclease mapping, poly(A)-RNA (20 μg) from Molt-4 cells was hybridized to a 5′ ^{32}P-labeled probe shown schematically in *C*. Lane 1, S1-resistant products; lane 2, undigested probe. (*B*) Mapping of the 5′ end of the α_4 gene by primer extension analysis. For primer extension reactions, poly(A)-RNA (20 μg) from Molt-4 cells was hybridized to the primer shown schematically in *C*. A sequencing reaction, using this same primer, was included to determine the exact sites of transcriptional initiation. The asterisk on the A (labeled +1) represents the major transcriptional start site, whereas asterisks at nucleotides −11 and +20 represents minor start sites. Size standards (STD) are the plasmid Bluescript SK digested with *Msp* I, and the sizes are given in nucleotides.

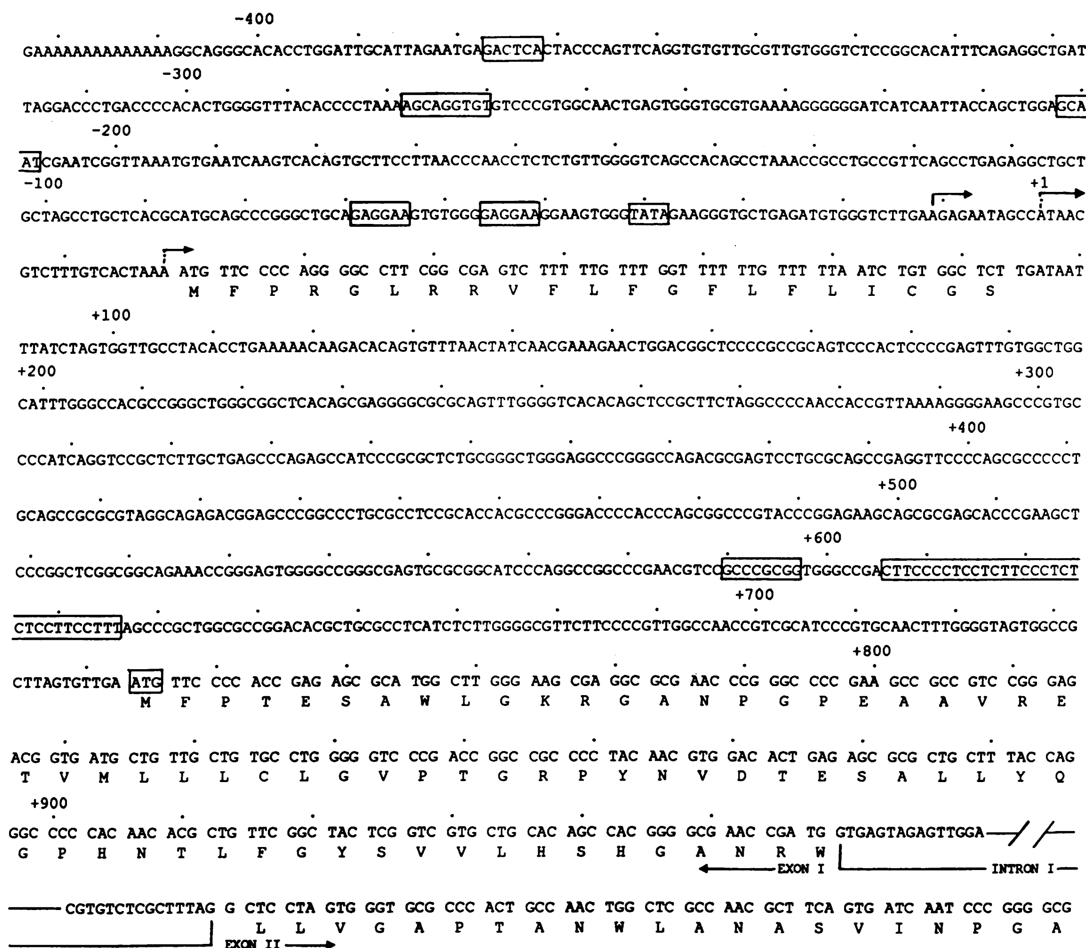

FIG. 4. DNA sequence of the 5′ end and flanking region of the α_4 gene. Arrows denote transcriptional start sites. The major start site is designated "+1." Boxed sequences are discussed in the text.

site (−274), and an AP-1 site (−377) (Fig. 4). The AP-1 site mediates transcriptional stimulation by transforming growth factor β_1 (26) and serum (27). α_4 expression is stimulated by transforming growth factor β_1 (28), and we have shown that α_4 mRNAs are increased in response to serum stimulation (data not shown). Therefore, an AP-1 site may be important in controlling expression of α_4. PU boxes have previously been identified as lymphoid- and myeloid-specific enhancers (29), suggesting that they could play a role in the tissue-specific expression of α_4. Intriguingly, α_4 is expressed in rhabdomyosarcomas, suggesting that the consensus binding site for the muscle-specific transcription factor MyoD (30) may be important for control of α_4 gene expression.

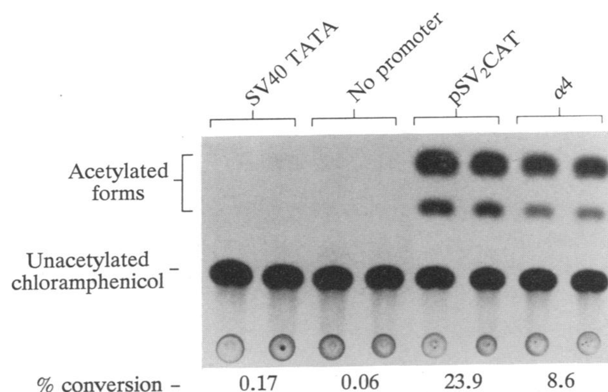

FIG. 5. Expression of an α_4–CAT fusion gene in transfection assays. An α_4–CAT fusion gene, containing ≈400 bp of α_4 gene 5′ flanking sequences, was transfected into the Jurkat cell line along with the plasmid pSV$_2$CAT, which contains the strong viral promoter from the simian virus 40 (SV40) early gene, a construct that contains only the simian virus 40 early gene TATA box (a weak promoter) fused to CAT gene, or a construct containing the CAT gene with no promoter.

An α_4–CAT fusion gene, containing ≈400 bp of 5′ flanking sequence, was transfected into the Jurkat cell line (Fig. 5). α_4 promoter activity was 143 times greater than with no promoter, 50 times greater than a weak promoter containing only the simian virus 40 early gene TATA box, and 30% of that observed with the strong simian virus 40 early gene promoter containing the enhancer sequences (pSV$_2$CAT).

Knowledge of α_4 gene promoter sequence should be valuable in the analysis of how expression of this multifunctional gene is controlled. Interestingly, the structure of the α_4 gene promoter is very different from that of the α_5 gene (T.M.B., E. Boedeker, E. Ruoslahti, S. Argraves, and D.C.D., unpublished results), which encodes another β_1 integrin FN receptor that is abundant on adherent cell types. Therefore, differences in tissue specificity between these two genes is probably dictated by alternate sets of promoter elements.

This work was supported by National Institutes of Health Grant HL43418 (D.C.D.). G.D.R. is a fellow of the American Lung Association and is supported by a Clinical Investigator Award from the National Heart, Lung, and Blood Institute (HL 02558). T.M.B. is supported by a Clinical Investigator Award from the National Heart, Lung, and Blood Institute (HL 02428), an American Lung Association Research Grant, and an RJR Nabisco Research Scholar Award.

1. Hynes, R. O. (1987) *Cell* **48,** 549–554.
2. Buck, C. A. & Horwitz, A. F. (1987) *Annu. Rev. Cell Biol.* **3,** 179–205.
3. Ruoslahti, E. & Pierschbacher, M. D. (1987) *Science* **238,** 491–497.
4. Hemler, M. E., Elices, M. J., Parker, C. & Takada, Y. (1990) *Immunol. Rev.* **114,** 45–65.
5. Wayner, E. A., Garcia-Pardo, A., Humphries, M. J., McDonald, J. A. & Carter, W. G. (1989) *J. Cell Biol.* **109,** 1321–1330.
6. Kornblihtt, A. R., Umezawa, K., Vibe-Petersen, K. & Barelle, F. (1985) *EMBO J.* **4,** 1755–1759.
7. Pytela, R., Pierschbacher, M. D. & Ruoslahti, E. (1985) *Cell* **40,** 191–198.
8. Argraves, S. W., Suzuki, S., Arai, H., Thompson, K., Pierschbacher, M. D. & Ruoslahti, E. (1987) *J. Cell Biol.* **105,** 1183–1190.
9. Osborn, L., Hession, C., Tizard, R., Vassallo, C., Luhowskyj, S., Chi-Rossa, G. & Lobb, R. (1989) *Cell* **59,** 1203–1211.
10. Elices, M. J., Osborn, L., Takada, Y., Crouse, C., Luhowskyj, S., Hemler, M. E. & Lobb, R. R. (1990) *Cell* **60,** 577–584.
11. Vartio, T., Laitinen, L., Naravanen, O., Cutolo, M., Thornell, L., Zardi, L. & Virtanen, I. (1987) *J. Cell Sci.* **88,** 419–430.
12. La Fleur, M., Beaulieu, A. D., Kreis, C. & Poubelle, P. (1987) *J. Biol. Chem.* **262,** 2111–2115.
13. Takada, Y., Elices, M. J., Crouse, C. & Hemler, M. E. (1989) *EMBO J.* **8,** 1361–1368.
14. Groux, H., Huet, S., Valentin, H., Pham, D. & Bernard, A. (1989) *Nature (London)* **339,** 152–154.
15. Holtzman, B. H. & Weissman, I. L. (1989) *EMBO J.* **8,** 1735–1741.
16. Holtzman, B. H., McIntyre, B. W. & Weissman, I. L. (1989) *Cell* **56,** 37–46.
17. Kozak, M. (1987) *Nucleic Acids Res.* **15,** 8125–8132.
18. Marth, J. D., Overell, R. W., Meier, K. E., Krebs, E. & Perlmutter, R. M. (1988) *Nature (London)* **332,** 171–173.
19. Dean, D. C., Blakeley, M. S., Newby, R. F., Ghazal, P., Hennighausen, L. & Bourgeois, S. (1989) *Mol. Cell. Biol.* **9,** 1498–1506.
20. Dean, D. C., Newby, R. F. & Bourgeois, S. (1988) *J. Cell Biol.* **106,** 2159–2170.
21. Dean, D. C., Bowlus, C. L. & Bourgeois, S. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 1876–1800.
22. Gubler, U. & Hoffman, B. J. (1983) *Gene* **25,** 283–287.
23. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
24. Southern, E. M. (1975) *J. Mol. Biol.* **98,** 503–511.
25. Saiki, R. K., Gelfand, D. H., Stoffel, S. J., Higuchi, R., Horn, G. T., Mullins, K. B. & Erlich, H. A. (1988) *Science* **488,** 487–491.
26. Kim, S. J., Angel, P., Lafyatis, R., Hattori, K., Sporn, M. B., Karin, M. & Roberts, A. B. (1990) *Mol. Cell. Biol.* **10,** 1492–1497.
27. Imler, J. L., Schatz, C., Wasylyk, C., Chatton, B. & Wasylyk, B. (1988) *Nature (London)* **332,** 275–278.
28. Heino, J., Ignotz, R. A., Hemler, M. E., Crouse, C. & Massague, J. (1989) *J. Biol. Chem.* **264,** 380–388.
29. Klemsz, M. J., McKercher, S. R., Celada, A., Van Beveren, C. & Maki, R. A. (1990) *Cell* **61,** 113–124.
30. Lassar, A. B., Buskin, J. N., Lockshon, D., Davis, R. L., Apone, S., Hauschka, S. D. & Weintraub, H. (1989) *Cell* **58,** 823–831.