Proc. Natl. Acad. Sci. USA
Vol. 91, pp. 4509–4513, May 1994
Biochemistry

Krüppel-associated boxes are potent transcriptional repression domains

(zinc finger protein/gene regulation/transcriptional regulation/GAL4)

JUDITH F. MARGOLIN*†, JOSH R. FRIEDMAN*, WOLFRAM K.-H. MEYER‡, HENRIK VISSING§, HANS-JÜRGEN THIESEN‡, AND FRANK J. RAUSCHER III*¶

*The Wistar Institute, Philadelphia, PA 19104; †Division of Oncology, Children's Hospital of Philadelphia, Philadelphia, PA 19104; ‡The Basel Institute for Immunology, Basel, Switzerland; and §Bioscience, Novo Nordisk Bagsvaerd, Denmark

Communicated by Herbert Weissbach, January 31, 1994

ABSTRACT **The Krüppel-associated box (KRAB) is a highly conserved, 75-aa region containing two predicted amphipathic α-helices. The KRAB domain is present in the amino-terminal regions of more than one-third of all Krüppel-class Cys_2His_2 zinc finger proteins and is conserved from yeast to man; however, its function is unknown. Here it is shown that the KRAB domain functions as a DNA binding-dependent transcriptional repressor when fused to a heterologous DNA-binding domain from the yeast GAL4 protein. A 45-aa segment containing one of the predicted KRAB amphipathic helices was necessary and sufficient for repression. Amino acid substitutions in the predicted helix abolished the repression function. These results assign a function, transcriptional repression, to the highly conserved KRAB box and define a minimal repression domain which may aid in identifying mechanisms of repression.**

An important theme which has emerged from the study of sequence-specific DNA-binding proteins is that functional domains contained in these polypeptides are very modular (1). A DNA-binding domain is usually fused to an "effector" (activation or repression) domain, which mediates a change in the rate of transcription initiation. Little is known about the structure or targets of the effector domains of transcription factors. The lack of any common amino acid sequence homology and/or structural motifs associated with effector domains has hampered their analysis. Activation domains are often rich in acidic amino acids and/or proline and glutamine (2, 3) and may contain amphipathic helices which form protein–protein interaction surfaces that contact the basal transcription machinery or intermediary proteins (4–6). However, some of the most potent activation domains (such as that of the herpes simplex virus-encoded protein VP16 and that of the yeast protein GAL4) may not require helicity or acidic character to function properly (7, 8).

Little is known of the nature and function of repression domains in transcription factors (9–11). Some transcriptional repressors contain regions that are rich in proline (12, 13), alanine (14–16), or serine/threonine (17) and may exhibit a net positive charge (18). In yeast, protein–protein interactions among transcription factors (19, 20) and altered chromatin structure (21, 22) may mediate repression. Our understanding of repression would be greatly enhanced by the identification of a conserved amino acid sequence motif common to repression domains. We have identified such a motif in Krüppel-class zinc finger proteins.

The Krüppel-associated box (KRAB) was identified as an ≈75-aa region of homology present in about one-third of the estimated 500 human zinc finger proteins (Fig. 1) (29) and is found exclusively in the amino termini of proteins that contain Krüppel-class Cys_2His_2 (C_2H_2) zinc fingers in their carboxyl termini (29). The KRAB domain is (*i*) subdivided into A and B boxes (Fig. 1), (*ii*) rich in charged amino acids, (*iii*) predicted to fold into two amphipathic helices and may contain a heptad repeat of methionine/leucine residues (see KOX1; Fig. 1), and (*iv*) present in all eukaryotes from yeast to man. Thus, the KRAB domain may have a highly conserved function in regulating transcription. We show here that four independently encoded KRAB domains function as repressors of transcription and define a minimal repression domain of 45 aa in the KRAB A box.

MATERIALS AND METHODS

Expression Vectors. The GAL4–KOX1, –ZNF133, –ZNF140, and –ZNF141 fusion genes were constructed by polymerase chain reaction (PCR) amplification of the KRAB domains using oligonucleotide primers. Each 5′ PCR primer contained an *Eco*RI site and each 3′ PCR primer contained a TGA stop codon and either an *Xba* I or an *Hin*dIII site. Standard PCRs were performed with the indicated zinc finger cDNAs (23, 24) as templates. Restriction endonuclease-digested PCR products were cloned into the pM1 vector (30). Each fusion protein contained GAL4-(1–147) (31), followed by proline, glutamic acid, phenylalanine, and the indicated segments of the KRAB-encoding cDNAs. The methionine encoded at nt 71 in the KOX1 cDNA (23) was designated as aa 1 for the KOX1-containing constructs. The ZNF cDNAs were fused to GAL4 as follows: ZNF133, nt 362–1037; ZNF141, nt 157–589; and ZNF140, nt 273–678. Amino acid substitutions in the KRAB domain were introduced by PCR-mediated, site-directed mutagenesis (32). All PCR-derived segments of DNA were sequenced on both strands. To construct KOX1-(1–161), a 5′ primer containing an *Hin*dIII site and a Kozak consensus sequence and a 3′ primer containing an *Xba* I site and a TGA stop codon were used for PCR. The product was cloned into the $pCB6^+$ expression vector, which contains the human cytomegalovirus immediate-early promoter (13). Expression of the KOX1-(1–161) protein was verified by immunoprecipitation with anti-KOX1 serum (J.F.M., unpublished data). Vectors expressing GAL4-E1A-(121–223) and GAL4-WT1-(1–298) proteins have been described (13, 33).

Cell Culture, Transfections, and Immunoprecipitation. NIH 3T3 mouse cells were grown in Dulbecco's modified Eagle's medium supplemented with 10% calf serum and were transfected by the calcium phosphate coprecipitation method (12, 13). COS-1 monkey cells were grown in Iscove's modified Dulbecco's medium supplemented with 10% fetal bovine

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: KRAB, Krüppel-associated box; CAT, chloramphenicol acetyltransferase; TK, thymidine kinase.

¶To whom reprint requests should be addressed at: The Wistar Institute, 3601 Spruce Street, Philadelphia, PA 19104.

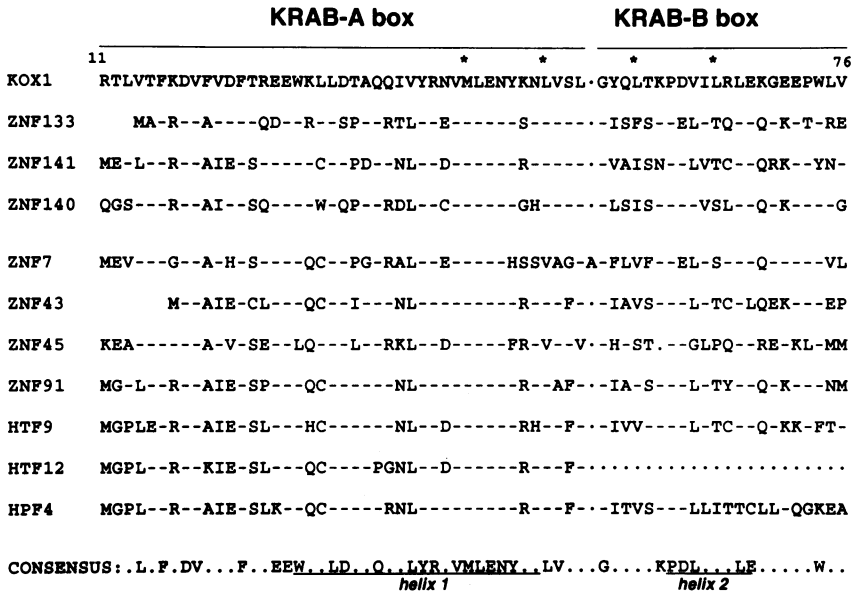

FIG. 1. Alignment of amino acid sequences of KRAB domains. The KRAB domains of 10 independent zinc finger gene products are aligned with the KOX1 sequence and the KRAB consensus (23). A dash indicates a residue identical to the corresponding residue in KOX1; a period indicates a gap introduced to facilitate alignment. The zinc finger regions of each protein are not shown, but the full-length amino acid sequences are known: KOX1 (23); ZNF 133 and ZNF 140 (H.V., unpublished data); ZNF133 (24); ZNF7 (25); ZNF43 (26); ZNF45 (27); ZNF91 (28); HTF9, HTF12, and HPF4 (29). The positions of the predicted α-helices are shown and the asterisks indicate the heptad repeat of hydrophobic residues (23).

serum and were transfected, labeled with [^{35}S]methionine, and subjected to immunoprecipitation using anti-GAL4 polyclonal serum (a gift from P. Silver, Harvard University) as described (12, 13, 34). The protocols for transfection, cell harvesting, and chloramphenicol acetyltransferase (CAT) assays have been described (12, 13). In brief, each 100-mm dish of NIH 3T3 cells received 5 μg of GAL–TK–CAT DNA and 5 μg of GAL4-fusion expression vector DNA (Fig. 2*A*), as well as 2 μg of the β-D-galactosidase-expressing plasmid DNA pON260 (35). Forty-eight hours after transfection, β-D-galactosidase and CAT activities were determined from cell extracts. The reaction products were quantitated with a Molecular Dynamics PhosphorImager and percent conversion values were calculated. Each transfection experiment was performed in duplicate from four to seven times. The percent conversion values shown represent the average of duplicates in a representative experiment. There were no toxic effects as a result of expressing the GAL4–KOX fusions in transient assays which could account for the repression observed.

RESULTS

Repression of Transcription Mediated by a GAL4–KOX1 Fusion Protein. The amino terminus of KOX1 contains KRAB A and B boxes in addition to 11 zinc fingers in the carboxyl terminus of the protein (23) (Fig. 2*A*). We fused the KRAB-containing domain, aa 1–161, to the DNA-binding domain of GAL4, aa 1–147, to create a fusion protein, GAL4–KOX1-(1–161). The GAL4-(1–147) protein has well-characterized DNA binding specificity (31), contains determinants for nuclear localization, and is neutral in transcriptional regulation when used in cotransfection assays (15). Each fusion protein was stably expressed in COS-1 cells (Fig. 2*B*). The expression vectors were cotransfected into NIH 3T3 cells with a CAT reporter vector (GAL–TK–CAT) which contained the TK basal promoter element fused to five synthetic GAL4 upstream activator sequences (12, 13, 34). The GAL–TK–CAT vector exhibited a high basal level of transcriptional activity when cotransfected with the expression vector containing the GAL4 DNA-binding domain alone (Fig. 2*C*). However, the transcriptional activity of the GAL–TK–CAT vector was strongly repressed (17-fold) when cotransfected with the GAL4–KOX1-(1–161) expression vector. As expected, the GAL4–WT1-(1–298) vector repressed and the GAL4–E1A-(121–223) vector activated the GAL–TK–CAT vector, as has been described (13, 33). Repression by GAL4–KOX1-(1–161) occurred at the level of transcription initiation (H.-J.T., unpublished data) and was dependent on the presence of GAL4 upstream activator sequences in the GAL–TK–CAT vector. A TK–CAT vector which lacked the GAL4 binding sites was unaffected by GAL4-KOX1-(1–161) (H.-J.T., unpublished data). Thus, like other repression domains (11–18), the amino terminus of KOX1, which contains the KRAB domain, can function as a DNA-binding-site-dependent repressor of transcription when fused to a heterologous DNA-binding domain.

To determine whether other KRAB domain-containing polypeptides could function as repression domains, we fused the amino termini of the KRAB-containing zinc finger proteins ZNF140, ZNF141, and ZNF133 to GAL4 (Fig. 3*A*). These proteins contain KRAB A and B boxes in their amino termini and variable numbers of zinc fingers in their carboxyl termini (24). Each GAL4–ZNF fusion protein was stably expressed in COS-1 cells (Fig. 3*B*) and each protein functioned as a potent repressor in a DNA-binding-site-dependent manner (Fig. 3*A*). Thus, a transferable transcriptional repression function appears to be a common activity of KRAB domain-containing proteins.

KRAB Boxes of KOX1 Are Sufficient for Repression. To localize the repression domain of KOX1, amino- and carboxyl-terminal truncations were made (Fig. 4). Each truncated protein was assayed for stable expression in COS-1 cells (unpublished data) and used in cotransfection assays with the GAL–TK–CAT reporter. Full repression activity was exhibited by GAL4–KOX1-(1–75). A further truncation eliminating the KRAB B box [GAL4–KOX1-(1–55)] partially inactivated repression. Use of amino-terminal truncations revealed that the first 11 aa were dispensable for function [GAL4–KOX1-(11–90)] but any deletion that extended into

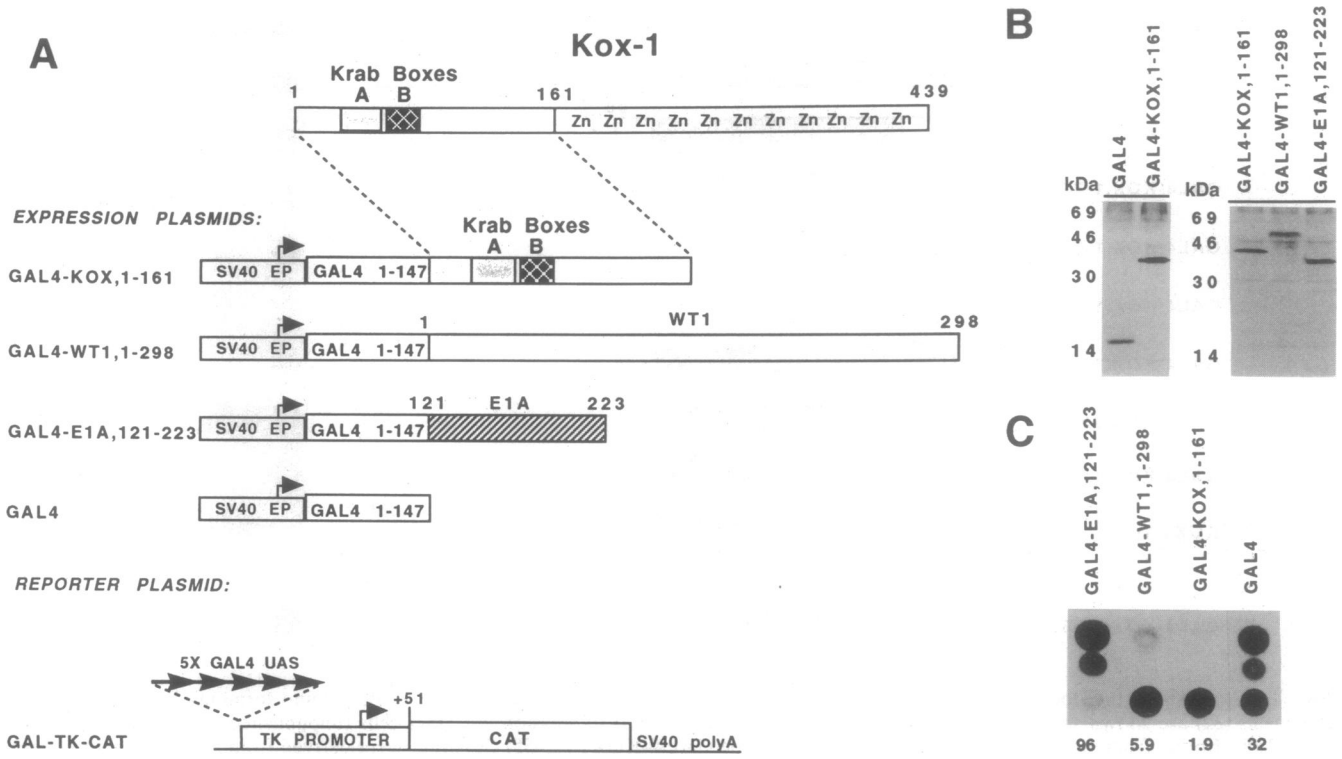

FIG. 2. Regulation of transcription by a GAL4–KOX1 fusion protein. (*A*) Expression vectors for use in transient-transfection assays contained the simian virus 40 (SV40) early promoter (EP) and polyadenylylation signals. The DNA-binding domain of GAL4, aa 1–147 (31), was fused to the indicated amino acids of KOX1, WT1, and E1A. The reporter plasmid (GAL–TK–CAT) contained five synthetic GAL4 upstream activation sequences (UAS), the thymidine kinase (TK) promoter, and the CAT gene (15). (*B*) The expression vectors were transfected into COS-1 cells and were analyzed by immunoprecipitation (34) with antiserum to GAL4-(1–147). The immunoprecipitated proteins were resolved by SDS/15% PAGE. (*C*) Calcium phosphate-mediated transfections were performed in murine NIH 3T3 fibroblasts. Each dish received GAL–TK–CAT (5 μg), the indicated expression plasmid (5 μg), and the β-D-galactosidase expression vector (2 μg). Forty-eight hours after transfection, CAT activity was measured in cell extracts that had been normalized for transfection efficiency. The numbers indicate average percent conversion of chloramphenicol in duplicate dishes derived from PhosphorImager analyses.

the KRAB A box abolished repression activity [GAL4–KOX1-(22–90)]. The GAL4–KOX1-(11–55) construct, which contained only 45 aa of the KRAB A box, showed significant repression activity. Thus, the KRAB A box of KOX1 is necessary and sufficient for transcriptional repression.

As a further control we constructed an expression vector [KOX1-(1–161)] that lacked the GAL4 DNA-binding domain but expressed the KOX1 amino terminus. This protein was properly expressed (unpublished data) but completely lacked repression activity (Fig. 4). Thus, the KRAB domain must be tethered to the DNA template in order to function as a repressor.

Identification of Amino Acids in the KRAB Domain That Are Important for Transcriptional Repression. To determine the amino acid sequence requirements within the KRAB domain necessary for transcriptional repression, we employed site-directed mutagenesis (32). We targeted the most highly conserved amino acids in the KRAB box consensus (Fig.

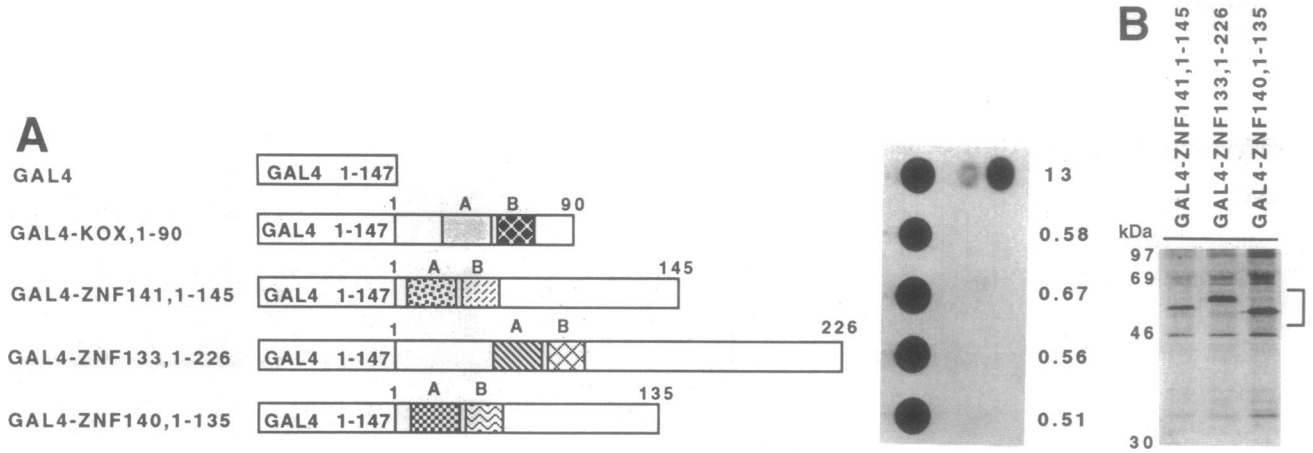

FIG. 3. KRAB domains from four different proteins function as repressors when fused to GAL4. The indicated amino-terminal amino acids from each ZNF protein (24) were fused to GAL4. Relative positions of the KRAB A and B boxes in the polypeptide segment are indicated by the textured boxes (*A*). The expression vectors were tested in COS-1 cells (*B*) and then cotransfected with the GAL–TK–CAT plasmid. Results of CAT assays (*A*) are expressed as in Fig. 2.

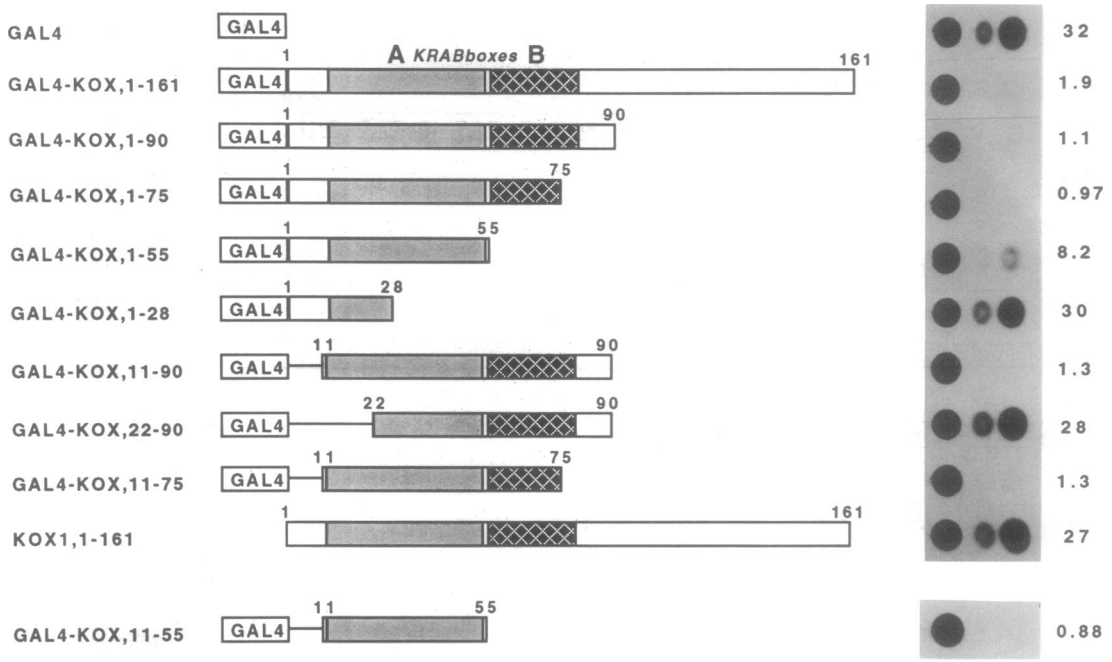

FIG. 4. Analysis of the domains of KOX1 required for repression. Truncations made in the KOX1 amino terminus by PCR-mediated mutagenesis (32) are shown schematically. Each protein was tested for stable expression in COS-1 cells (unpublished data) and its ability to repress transcription from the GAL–TK–CAT vector. Data for the GAL4–KOX1-(11–55) construct were derived from a different experiment than the one shown for all the other truncations. The GAL4 control for this experiment (13% conversion) is shown in Fig. 3*A*.

5*A*). After analysis of expression in COS-1 cells (Fig. 5*B*), the plasmids expressing the mutated KRAB-domain proteins were utilized in cotransfection assays (Fig. 5*A*). Substitutions in residues KP and LE in the B box had little effect on repression, which was consistent with our observation that the KRAB B domain could be deleted without loss of repression. However, substitution in the A box at the DV, EEW, and MLE sequences significantly inhibited the ability of GAL4–KOX1-(1–90) to repress transcription. The leucine residue in the MLE sequence occurs in the context of the highly conserved VMLENY motif common to almost all KRAB A boxes (Fig. 1) and is one of the heptad repeats of leucines ($MX_6LX_6LX_6L$, where X is any amino acid) capable of forming an amphipathic helix. This sequence may form a potential protein–protein interaction surface important for repression of transcription. Thus, we have identified discrete mutations in the KRAB domain that inactivate the transcriptional repression.

DISCUSSION

These studies have assigned a biochemical function, transcriptional repression, to the KRAB domain estimated to be present in about one-third of all human zinc finger genes (29). The minimal repression domain in KOX1 is a 45-aa segment in the A box which, when fused to GAL4-(1–147), converted this transcriptionally neutral DNA-binding domain into a potent repressor. Thus, like other repression domains that function when fused to heterologous DNA-binding domains (11–18, 36), the KRAB box is a transferable, DNA-binding-site-dependent repressor of transcription. We have also shown that the GAL4–KRAB protein can (*i*) repress transcription when binding sites are placed 2–3 kb downstream (3′) to the initiation site, and (*ii*) repress both human immunodeficiency virus Tat-activated or Sp1-activated transcription (H.-J.T. and F.J.R., unpublished data). Unlike other alanine- or proline-rich repression domains, the KRAB domain is present in a highly conserved family of proteins

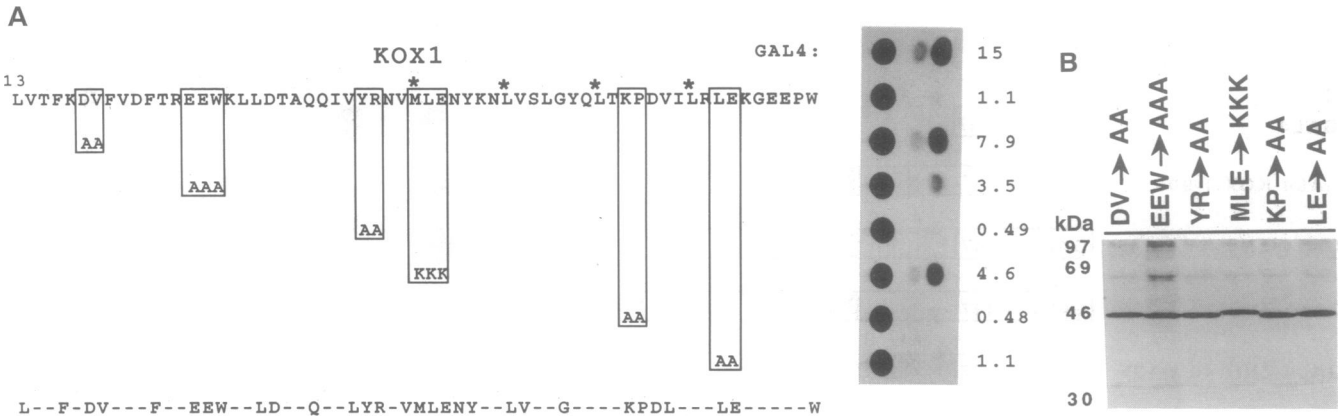

FIG. 5. Amino acid substitutions in conserved regions of the KOX1 KRAB domain inhibit transcriptional repression. The boxed amino acids in GAL4–KOX1-(1–90) were mutated to the indicated amino acids (alanine or lysine) by PCR-mediated mutagenesis. After DNA sequence analysis and verification of stable expression in COS-1 cells (*B*), the GAL4–KOX1 fusions were tested for ability to repress transcription from the GAL–TK–CAT vector. Results of CAT assays are given as in Fig. 2.

sharing a common primary amino acid sequence and a predicted secondary structure. These characteristics suggest that the KRAB domains in many different proteins may share a common intracellular target, which may simplify identification of proteins which interact with the KRAB domain *in vivo*.

It is intriguing that the KRAB domain is only found in the amino termini of proteins that have Krüppel-type C_2H_2 zinc fingers in their carboxyl termini. In contrast, other effector domains such as acidic activation or proline-rich regions are found in transcription factors with very diverse types of DNA-binding domains. The strong association of KRAB boxes with Krüppel-type zinc fingers suggests that the two domains may function in concert. It will be important to determine the transcriptional regulation potential of a KRAB domain-containing protein when bound to its proper DNA recognition sequence via the zinc finger region. However, currently there is no recognition sequence known for any KRAB domain-containing transcription factor.

The mechanism(s) of repression by the KRAB domain is unclear. The KRAB domain must be tethered to the template DNA via fusion to a DNA-binding domain, since the KRAB domain alone did not repress transcription. Thus, it does not appear that nonspecific "squelching" or titration of basal factors is occurring. The presence of a charged, amphipathic helix with potential leucine zipper-like structure in the A box is suggestive of a protein–protein interaction domain. When bound to DNA, the KRAB domain may interact with elements of the basal transcription machinery, the activation domains of other transcription factors, or their coactivator/adaptor proteins. The amino acid substitutions in the KRAB domain which inactivate its function should prove useful in identifying and distinguishing protein–protein interactions required for repression.

We thank S. Madden for the GAL4–WT1-(1–298) and KOX1-(1–161) constructs; J. Morris for help in immunoprecipitation analyses; P. Silver for GAL4 antiserum; I. Sadowski for GAL4 expression vectors; Y. Shi for the GAL–TK–CAT plasmid; N. Galili, T. Halazonetis, and G. Rovera for reviewing the manuscript; and M. Marinelli for preparing the manuscript. This work was supported by National Institutes of Health Grants CA52009, CA47983, and CA10815 and the Hansen and the Mary A. H. Rumsey Foundations (F.J.R.) and by the Molecular Approaches to Pediatric Science (MAPS-HD28815) and the American Society of Clinical Oncology 1992 Young Investigator fellowships (J.F.M.). The Basel Institute for Immunology was founded and is supported by F. Hoffman–La Roche and Company, Ltd., Basel. F.J.R. is a Pew Scholar in the Biomedical Sciences.

1. Pabo, C. O. & Sauer, R. T. (1992) *Annu. Rev. Biochem.* **61,** 1053–1095.
2. Mitchell, P. J. & Tjian, R. (1989) *Science* **245,** 371–378.
3. Johnson, P. F. & McKnight, S. L. (1989) *Annu. Rev. Biochem.* **58,** 799–839.
4. Stringer, K. F., Ingles, C. J. & Greenblatt, J. (1990) *Nature (London)* **345,** 783–786.
5. Berger, S. L., Cress, W. D., Cress, A., Triezenberg, S. J. & Guarente, L. (1990) *Cell* **61,** 1199–1208.
6. Berger, S. L., Pina, B., Silverman, N., Marcus, G. A., Agapite, J., Regier, J. L., Triezenberg, S. J. & Guarente, L. (1992) *Cell* **70,** 251–265.
7. Cress, W. D. & Triezenberg, S. J. (1991) *Science* **251,** 87–90.
8. Leuther, K. K., Salmeron, J. M. & Johnston, S. A. (1993) *Cell* **72,** 575–585.
9. Levine, M. & Manley, J. L. (1989) *Cell* **59,** 405–408.
10. Renkawitz, R. (1990) *Trends Genet.* **6,** 192–196.
11. Cowell, I. G., Skinner, A. & Hurst, H. C. (1992) *Mol. Cell. Biol.* **12,** 3070–3077.
12. Madden, S. L., Cook, D. M., Morris, J. F., Gashler, A., Sukhatme, V. P. & Rauscher, F. J., III (1991) *Science* **253,** 1550–1553.
13. Madden, S. L., Cook, D. M. & Rauscher, F. J., III (1993) *Oncogene* **8,** 1713–1720.
14. Licht, J. D., Grossel, M. J., Figge, J. & Hansen, U. M. (1990) *Nature (London)* **346,** 76–79.
15. Shi, Y., Seto, E., Chang, L. S. & Shenk, T. (1991) *Cell* **67,** 377–388.
16. Han, K. & Manley, J. L. (1993) *EMBO J.* **12,** 2723–2733.
17. Gashler, A. L., Swaminathan, S. & Sukhatme, V. P. (1993) *Mol. Cell. Biol.* **13,** 4556–4571.
18. Saha, S., Brickman, J. M., Lehming, N. & Ptashne, M. (1993) *Nature (London)* **363,** 648–652.
19. Keleher, C. A., Goutte, C. & Johnson, A. D. (1988) *Cell* **53,** 927–936.
20. Keleher, C. A., Redd, M. J., Schultz, J., Carlson, M. & Johnson, A. D. (1992) *Cell* **68,** 709–719.
21. Shimizu, M., Roth, S. Y., Szent-Gyorgyi, C. & Simpson, R. T. (1991) *EMBO J.* **10,** 3033–3041.
22. Roth, S. Y., Shimizu, M., Johnson, L., Grunstein, M. & Simpson, R. T. (1992) *Genes Dev.* **6,** 411–425.
23. Thiesen, H. J. (1990) *New Biol.* **2,** 363–374.
24. Tommerup, N., Aagaard, L., Boel, E., Baxendale, S., Bates, G., Lehrach, H. & Vissing, H. (1993) *Hum. Mol. Genet.* **2,** 1571–1575.
25. Lania, L., Donti, E., Pannuti, A., Pascucci, A., Pengue, G., Feliciello, I., La Mantia, G., Lanfrancone, L. S. & Pelicci, P. G. (1990) *Genomics* **6,** 333–340.
26. Lovering, R. & Trowsdale, J. (1991) *Nucleic Acids Res.* **19,** 2921–2928.
27. Constantinou-Deltas, C. D., Gilbert, J., Bartlett, R. J., Herbstreith, M., Roses, A. D. & Lee, J. E. (1992) *Genomics* **12,** 581–589.
28. Bellefroid, E. J., Marine, J. C., Reid, T., Lecocq, P. J., Riviére, M., Amemiya, A., Poncelet, D. A., Coulie, P. G., de Jong, P., Szpirer, C., Ward, D. C. & Martial, J. A. (1993) *EMBO J.* **12,** 1363–1374.
29. Bellefroid, E. J., Poncelet, D. A., Lecocq, P. J., Revelant, O. & Martial, J. A. (1991) *Proc. Natl. Acad. Sci. USA* **88,** 3608–3612.
30. Sadowski, I., Bell, B., Broad, P. & Hollis, M. (1992) *Gene* **118,** 137–141.
31. Ma, J. & Ptashne, M. (1987) *Cell* **48,** 847–853.
32. Ho, S. N., Hunt, H. D., Horton, R. M., Pullen, J. K. & Pease, L. R. (1989) *Gene* **77,** 51–59.
33. Lillie, J. W. & Green, M. R. (1989) *Nature (London)* **338,** 39–44.
34. Morris, J. F., Madden, S. L., Tournay, O. E., Cook, D. M., Sukhatme, V. P. & Rauscher, F. J., III (1991) *Oncogene* **6,** 2339–2348.
35. Spaete, R. R. & Mocarski, E. S. (1985) *J. Virol.* **56,** 135–143.
36. Witzgall, R., O'Leary, E., Gessner, R., Ouellette, A. J. & Bonventre, J. V. (1993) *Mol. Cell. Biol.* **13,** 1933–1942.