



Published in final edited form as:

J Biol Chem. 1997 July 18; 272(29): 18504–18507.

Two Potent Nuclear Localization Signals in the Gut-enriched Krüppel-like Factor Define a Subfamily of Closely Related Krüppel Proteins*

Janiel M. Shields^{†,§} and Vincent W. Yang^{†,¶,||}

‡ Department of Medicine, The Johns Hopkins University School of Medicine, Baltimore, Maryland 21205

¶ Department of Biological Chemistry, The Johns Hopkins University School of Medicine, Baltimore, Maryland 21205

Abstract

The gut-enriched Krüppel-like factor (GKLF) is a newly identified transcription factor that contains three C₂H₂ Krüppel-type zinc fingers. Previous immunocytochemical studies indicate that GKLF is exclusively localized to the nucleus. To identify the nuclear localization signal (NLS) within GKLF, cDNA constructs with various deletions in the coding region of GKLF were generated and analyzed by indirect immunofluorescence in transfected COS-1 cells. In addition, constructs fusing regions representing putative NLSs of GKLF to green fluorescent protein (GFP) were generated and examined by fluorescence microscopy in similarly transfected cells. The results indicate that GKLF contains two potent, independent NLSs: one within the zinc fingers and the other in a cluster of basic amino acids (called 5' basic region) immediately preceding the first zinc finger. In comparison, putative NLSs within the zinc fingers and the 5' basic region of a related Krüppel protein, zif268/Egr-1, are relatively less efficient in their ability to translocate GFP into the nucleus. A search in the protein sequence data base revealed that despite the existence of numerous Krüppel proteins, only two, the lung Krüppel-like factor (LKLF) and the erythroid Krüppel-like factor (EKLF), exhibit similar NLSs to those of GKLF. These findings indicate that GKLF, LKLF, and EKLF are members of a subfamily of closely related Krüppel proteins.

Various mechanisms that are responsible for nuclear localization of eukaryotic transcription factors have been proposed. Most transcription factors contain one or more nuclear localization signal (NLS),¹ which, when recognized by nuclear transport proteins, results in the translocation of the transcription factor to the nuclear pore complex. Subsequent translocation across the nuclear membrane occurs in an ATP-dependent fashion (1). By inspecting the amino acid sequences of a large number of transcription factors, two types of NLSs have been defined (2,3). The first type, called a “core” NLS, contains four or more arginine and lysine residues within a hexapeptide and is frequently flanked by acidic residues or “helix-breakers” such as proline and glycine (3). The second type of NLS is “bipartite” and consists of two clusters of basic amino acids separated by a short nonbasic peptide. It is hypothesized that the two clusters of basic amino acids in a bipartite NLS are brought to a juxtaposed position due to protein folding and are subsequently recognized by the nuclear import machinery (2). In an analysis

*This work was supported by Grants DK44484 and DK52230 from the National Institutes of Health (to V. W. Y.).

|| To whom correspondence should be addressed: Dept. of Medicine, GI Division, Ross 918, The Johns Hopkins University School of Medicine, 720 Rutland Ave., Baltimore, MD 21205. Tel.: 410-955-9691; Fax: 410-955-9677; E-mail: vyang@welchlink.welch.jhu.edu.

§Recipient of an National Research Service Award from the National Institutes of Health.

¹The abbreviations used are: NLS, nuclear localization signal; BKLF, basic Krüppel-like factor; BTEB2, basic transcription element binding protein 2; EKLF, erythroid Krüppel-like factor; EZF, epithelial zinc finger; GFP, green fluorescent protein; GKLF, gut-enriched Krüppel-like factor; LKLF, lung Krüppel-like factor; aa, amino acid(s).

of the sequences of 117 transcription factors, 106 were found to contain one or more core NLS, whereas relatively few contain a bipartite NLS (2,3). Interestingly, many putative NLSs are present in close proximity to the DNA-binding domains of transcription factors, exemplified by the bZIP proteins c-Fos and c-Jun, and the bHLH proteins Myc, Max, and Myo D1 (3). This conserved arrangement seems to suggest that DNA-binding motifs and nuclear localization signals may have coevolved.

We recently identified a novel transcription factor named gut-enriched Krüppel-like factor (GKLF) which contains three C₂H₂ Krüppel-type zinc fingers (5). Expression of *GKLF* is enriched in the intestinal tract with the highest level of transcript found in the post-mitotic epithelial cells of the colon. *In vitro*, expression of *GKLF* is increased in culture conditions that induce growth arrest such as serum deprivation or contact inhibition. Furthermore, enforced expression of *GKLF* in transfected cells results in the inhibition of DNA synthesis. Together, these results indicate that *GKLF* is a growth arrest-associated, epithelial-specific gene. Subsequently, GKLF was independently identified by another group, which named it epithelial zinc finger (EZF), and shown also to be expressed at high levels in the epidermal layers of the skin (6). These findings suggest that GKLF/EZF may be involved in growth regulation and perhaps terminal differentiation of specific epithelial tissues. The primary amino acid sequence in the zinc finger region of GKLF exhibits a high degree of identity with several previously identified Krüppel proteins, including lung Krüppel-like factor (LKLF (7)), erythroid Krüppel-like factor (EKLF (8)), and basic transcription element-binding protein 2 (BTEB2 (9)). Because of the highly homologous nature of the zinc finger sequences of LKLF, EKLF, and BTEB2, it has been proposed that the three belong to the same multigene family (7).

Our previous studies showed that GKLF localized exclusively to the nucleus of cells transfected with a GKLF-expressing plasmid construct (5). To further investigate the structure-function relationship of GKLF with regard to nuclear localization, we determined its NLS in the present study. We show that GKLF contains two potent NLSs, each of which is sufficient to direct GKLF or an unrelated polypeptide into the nucleus. One of the NLSs resides in the zinc fingers and the other in a region (called 5' basic region) immediately amino-terminal to the first zinc finger. In contrast, by our studies and previous reports, nuclear localization of a related Krüppel protein, zif268/Egr-1, appears to require the participation of both the 5' basic region and the zinc fingers (4,10). Our results suggest that the Krüppel family of transcription factors can further be divided into subfamilies based on the sequences required for nuclear localization.

EXPERIMENTAL PROCEDURES

DNA Constructs

A GKLF cDNA containing the entire 483 amino acid (aa) open reading frame cloned into the mammalian expression vector, PMT3 (11), was described previously (5). Three mutant constructs with progressive deletions from the 3' end of the GKLF coding region were generated by digesting the full-length cDNA with appropriate restriction endonucleases (Fig. 1*a*). PMT3-GKLF-(1–441) contains the 5' basic region (broadly defined as the 20 amino acids (residues 382–401) immediately preceding the first cysteine of the first zinc finger of GKLF) and a deletion of the carboxyl-terminal 1½ zinc fingers. PMT3-GKLF-(1–401) contains the 5' basic region and a deletion of all three zinc fingers. PMT3-GKLF-(1–349) contains a deletion of both the 5' basic region and the zinc fingers. In addition, a construct containing only the 5' basic region and the three zinc fingers of GKLF was generated (PMT3-GKLF-(350–483)).

All green fluorescent protein (GFP) fusion proteins were generated in the expression vector, pEGFP-C3 (Clontech Laboratories, Inc.). cDNAs corresponding to peptides of those shown in Fig. 2*a* were generated by the polymerase chain reaction using appropriate primers and fused

to the carboxyl terminus of GFP. All constructs were sequenced to ensure the accuracy of the reading frames and to verify the fidelity of the polymerase chain reaction.

Transfection and Immunocytochemistry

Transient transfections were performed in COS-1 cells by lipofection (Life Technology, Inc.) as described previously (5). The procedure for indirect immunofluorescence analysis of GKLF in transfected cells using a primary polyclonal rabbit antiserum directed against GKLF and fluorescein isothiocyanate-conjugated secondary goat anti-rabbit serum has also been described (5). For visualization of cells transfected with the GFP fusion constructs, cells were fixed and permeabilized in an identical manner to those described for indirect immunofluorescence (5) and visualized with a Zeiss Axioskop 20 microscope equipped for epifluorescence.

RESULTS

Nuclear localization of GKLF was first examined by indirect immunofluorescence of COS-1 cells that had been transiently transfected with the full-length or various deletion constructs of GKLF in PMT3. Fig. 1 shows the results of one such experiment, which is representative of three independent experiments performed. The expressed GKLF protein was found to be present in the nucleus of cells transfected with constructs that retained the 5' basic region of GKLF (*constructs A, B, C, and E* (Fig. 1)). In contrast, deletion of the 5' basic region resulted in a significant distribution of the protein to the cytoplasm (*construct D* (Fig. 1)). Cells transfected with the empty PMT3 vector showed only minimal background staining (*construct F* (Fig. 1)). These results indicate that the 5' basic region of GKLF is both necessary and sufficient for nuclear localization.

To further delineate the NLS of GKLF, DNA constructs fusing various regions of GKLF to the carboxyl terminus of GFP were generated and analyzed by fluorescence microscopy in transiently transfected COS-1 cells. As seen in Fig. 2, GFP alone was localized throughout the cell (*construct F* (Fig. 2)). In contrast, the three zinc fingers of GKLF, devoid of the 5' basic region, were able to redistribute the GFP fusion protein exclusively to the nucleus (*construct A* (Fig. 2)). Moreover, not all three zinc fingers were required for nuclear localization, since a construct retaining only the amino-terminal 1½ fingers also localized to the nucleus (*construct B* (Fig. 2)). This latter observation is different from that of a previous study involving a GKLF-related protein, zif268/Egr-1, in which deletion of any of its zinc fingers resulted in a loss of nuclear localization (10). Finally, a construct containing only the 5' basic region of GKLF was also able to drive the GFP fusion protein into the nucleus (*construct C* (Fig. 2)). These results indicate that the 5' basic region as well as the zinc fingers of GKLF function as potent NLSs and that each is capable of independently translocating a heterologous protein into the nucleus.

The NLS of zif268/Egr-1 has been examined in detail in two previous studies (4,10). While one study suggests that the three zinc fingers of zif268/Egr-1 are sufficient for nuclear localization (10), another study indicates that the 5' basic region of zif268/Egr-1 in combination with its zinc fingers are necessary for full localization to the nucleus (4). Because the 5' basic region of GKLF alone is a sufficient and strong NLS, we compared the ability of this region of zif268/Egr-1 to that of GKLF to localize GFP to the nucleus. Our results confirmed the previous observations (4,10) that the 5' basic region of zif268/Egr-1 functions as an NLS (*construct D* (Fig. 2)). However, the potency of this region to localize GFP to the nucleus appears to be relatively lower than that of the corresponding region of GKLF since in cells transfected with construct D, nuclear fluorescence was relatively weak and cytoplasmic fluorescence could be seen in a fair number of transfected cells when compared with cells transfected with construct C (Fig. 2).

Last, it was previously shown that a point mutation converting an arginine to a glycine in the third zinc finger of zif268/Egr-1, a region with a core NLS sequence, destroyed its nuclear localization (10). Thus, we analyzed the ability of this putative NLS to direct nuclear localization by the GFP fusion approach. Surprisingly, the results show that despite the presence of a core NLS in this region, the fusion protein was only weakly localized to the nucleus (*construct E* (Fig. 2)). Combining the results of our study and the two previous reports, it appears that the relative nuclear localizing activity of the 5' basic region and the zinc fingers of GKLf are stronger than that of the corresponding regions of zif268/Egr-1 (Figs. 1 and 2; Refs. 4 and 10).

DISCUSSION

Protein nuclear localization is a relatively new topic in the field of protein transport. In the last decade, significant progress has been made toward the understanding of the mechanisms that mediate localization of proteins to the nucleus. This is in part due to the availability of data bases containing amino acid sequences of a large number of transcription factors. By comparing these sequences, it becomes clear that most transcription factors depend on specific nuclear localization signals to achieve efficient translocation into the nucleus (2,3). Further investigation of the function of individual NLSs should lead to a better understanding of the mechanisms responsible for nuclear localization. In addition, it is becoming clear that many transcription factors contain more than one NLS and that the rate of nuclear import may be directly related to the number of NLSs present (22). Thus, the process of nuclear localization may reflect yet another level of regulation in transcription factors.

The goal of the present study was to delineate the NLS within a newly identified zinc finger-containing transcription factor, GKLf (5), also known as EZF (6). The results of our study clearly demonstrate that GKLf contains two NLSs, each capable of functioning independently and efficiently to translocate either GKLf or a heterologous protein into the nucleus. One of these NLSs resides in the 5' basic region of GKLf, which includes a core NLS sequence (four arginines and lysines within a hexapeptide) from aa residues 385–390 (PKRGRR). The second NLS is located within the zinc finger portion of GKLf within the amino-terminal 1½ zinc finger region, which alone is sufficient to confer nuclear localization.

The finding that the zinc fingers of GKLf contain an NLS is both surprising and interesting, since no putative NLS (core or bipartite) is found within the zinc finger region. Nonetheless, our results are consistent with the conclusion from a previous study that a “global” structure of zinc fingers, rather than specific sequences, serves as an NLS for zif268/Egr-1 (10). A notable difference between GKLf and zif268/Egr-1 is that the latter requires the participation of all three zinc fingers for efficient nuclear translocation while GKLf requires only the first 1½ zinc fingers. It appears that while both these proteins belong to the Krüppel family of transcription factors due to conservation of their zinc finger sequences, they appear to have diverged sufficiently that their signals for nuclear localization are structurally different.

A comparison of GKLf's aa sequence with those stored in the GenBank™ data base revealed several transcription factors with highly homologous sequences to the zinc finger region of GKLf. These proteins include LKLf, EKLF, and BTEB2 (5). In fact, before the publication of our study on the identification of GKLf, Lingrel and colleagues proposed that LKLf, EKLF, and BTEB2 belong to the same multigene family (7). Indeed, the percent amino acid identity between the zinc finger regions of GKLf and LKLf, EKLF, and BTEB2, is 91, 85, and 82%, respectively. If instead, the 20 amino acids within the 5' basic region of these proteins are compared, GKLf and LKLf are 90% identical while GKLf and EKLF are 65% identical (Fig. 3). More importantly, the 5' basic region of GKLf contains an identical core NLS to that of LKLf (PKRGRR), which is nearly identical to that of EKLF (SKRGRR). Since the 5' basic

region of GKLf was shown to function as a potent NLS, and since GKLf is highly similar to both LKLf and EKLF in the corresponding region, we would predict that the 5' basic region of LKLf and EKLF would also function as a strong NLS.

The one exception to the hypothesis proposed by Lingrel and colleagues (7) seems to be BTEB2. Despite an overall similarity in the aa sequence between the zinc fingers of GKLf and BTEB2 (82%), the sequences in the 5' basic region of the two proteins are very different, sharing only 15% identity (Fig. 3). In fact, when other Krüppel proteins with conserved zinc finger sequences are compared, the 5' basic region of BTEB2 is more related to a different group of proteins, which include BKLF, CPBP, and SP1 (Fig. 3). The 5' basic region of two of these proteins, BTEB2 and SP1, do not even contain a core NLS. The aa sequences of the 5' basic region of another group of zinc finger proteins, including early growth response α , transforming factor β -inducible early gene, and GC box-binding protein, are even more divergent from those of the GKLf family of proteins (Fig. 3). In fact, no sequences that even resemble an NLS can be identified in this group. Taken together, our study suggests that the Krüppel family of transcription factors can be divided into subfamilies based on homology of the 5' basic region, a region clearly shown to be important for the nuclear localization of GKLf. Our study also demonstrates that GKLf, LKLf, and EKLF are indeed closely related members of the same subfamily whereas BTEB2 belongs to a different subfamily.

Our study indicates that the 5' basic region of zif268/Egr-1 contains an NLS, which functionally does not appear to be as strong as that of GKLf. This result is consistent with those of two previous studies. In one study (4), the 5' basic region of zif268/Egr-1 was fused to β -galactosidase, and in another study (10), the region was retained in constructs in which all three zinc fingers of zif268/Egr-1 were deleted. In each case, an incomplete nuclear localization was observed. An inspection of the aa sequence in the 5' basic region of zif268/Egr-1 and other related proteins such as Egr-2 and Egr-3 (Fig. 3) showed that they do not contain a core NLS but exhibit characteristics of a bipartite NLS. It is possible that secondary or tertiary structure may contribute to the function of this region as an NLS. More likely, however, is that this region contributes to the overall nuclear localization of zif268/Egr-1 when it is associated with the protein's zinc fingers as suggested previously (4,10).

Another interesting finding derived from the analysis of the zif268/Egr-1's NLS is the effect of the carboxyl half of its third zinc finger in translocating GFP to the nucleus (*construct E* (Fig. 2)). Previous studies indicate that a mutation of the first arginine residue in this region in the context of the whole protein destroyed nuclear localization despite the fact that the zinc finger structure was maintained (10). These results suggest that this peptide sequence should be a potent NLS. Indeed, a core NLS sequence is present in this region, although it appears in an infrequently observed pattern where an arginine residue is separated from a lysine residue by two nonbasic aa residues (RKRHTK). In the study by Boulikas (3), NLSs with this type of configuration account for only 5% of 271 core NLSs examined. When determined empirically, this particular motif was shown to be relatively inefficient in directing a fused albumin protein into the nucleus (21). This result is consistent with our finding that this region by itself confers fairly poor nuclear localization (*construct E* (Fig. 2)).

In conclusion, we have shown that GKLf contains two potent and independent nuclear localization signals, the sequences of which are highly conserved in two other Krüppel-like factors, LKLf and EKLF. In addition, by sequence and/or structural analysis of the 5' basic regions of other Krüppel proteins, three additional subfamilies are identified, each predicted to utilize this region to a somewhat different extent for nuclear localization. These differences allow the separation of the various Krüppel proteins into distinct subfamilies and may reflect differences in the mechanisms regulating nuclear import among the subfamilies.

Acknowledgements

We thank the Genetics Institute for providing the PMT3 plasmid.

References

1. Richardson WD, Mills AD, Dilworth SM, Laskey RA, Dingwall C. *Cell* 1988;50:655–664. [PubMed: 3125984]
2. Boulikas T. *Crit Rev Eukaryotic Gene Exp* 1993;3:193–227.
3. Boulikas T. *J Cell Biol* 1994;55:32–58.
4. Gashler AL, Swaminathan S, Sukhatme VP. *Mol Cell Biol* 1993;13:4556–4571. [PubMed: 8336701]
5. Shields JM, Christy RJ, Yang VW. *J Biol Chem* 1996;271:20009–20017. [PubMed: 8702718]
6. Garrett-Sinha LA, Eberspaecher H, Seldin MF, de Crombrughe B. *J Biol Chem* 1996;271:31384–31390. [PubMed: 8940147]
7. Anderson KP, Kern CB, Crable SS, Lingrel JB. *Mol Cell Biol* 1995;15:5957–5965. [PubMed: 7565748]
8. Miller IJ, Bieker JJ. *Mol Cell Biol* 1993;13:2776–2786. [PubMed: 7682653]
9. Sogawa K, Imataka H, Yamasaki Y, Kusume H, Abe H, Fujii-Kuriyama Y. *Nucleic Acids Res* 1993;21:1527–1532. [PubMed: 8479902]
10. Matheny C, Day ML, Milbrandt J. *J Biol Chem* 1994;269:8176–8181. [PubMed: 8132543]
11. Swick AG, Janicot M, Cheneval-Kastelic T, McLenithan JC, Lane MD. *Proc Natl Acad Sci U S A* 1992;89:1812–1816. [PubMed: 1542676]
12. Crossley M, Whitelaw E, Perkins A, Williams G, Fujiwara Y, Orkin SH. *Mol Cell Biol* 1996;16:1695–1705. [PubMed: 8657145]
13. Koritschoner NP, Bocco JL, Panzetta-Dutari GM, Dumur CI, Flury A, Patrino LC. *J Biol Chem* 1997;272:9573–9580. [PubMed: 9083102]
14. Kadonaga JT, Carner KR, Masiarz FR, Tjian R. *Cell* 1987;51:1079–1090. [PubMed: 3319186]
15. Blok LJ, Grossmann ME, Perry JE, Tindall DJ. *Mol Endocrinol* 1995;9:1610–1620. [PubMed: 8584037]
16. Subramaniam M, Harris SA, Oursler MJ, Rasmussen K, Riggs BL, Spelsberg TC. *Nucleic Acids Res* 1995;23:4907–4912. [PubMed: 8532536]
17. Christy BA, Lau LF, Nathans D. *Proc Natl Acad Sci U S A* 1988;85:7857–7861. [PubMed: 3141919]
18. Sukhatme VP, Cao XM, Chang LC, Tsai-Morris CH, Stamenkovich D, Ferreira PC, Cohen DR, Edwards SA, Shows TB, Curran T, Le Beau MM, Adamson ED. *Cell* 1988;53:37–43. [PubMed: 3127059]
19. Lemaire P, Revelant O, Bravo R, Charnay P. *Proc Natl Acad Sci U S A* 1988;85:4691–4695. [PubMed: 3133658]
20. Patwardhan S, Gashler A, Siegel MG, Chang LC, Joseph LJ, Shows TB, Le Beau MM, Sukhatme VP. *Oncogene* 1991;6:917–928. [PubMed: 1906159]
21. Goldfarb DS, Garipey J, Schoolnik G, Kornberg RD. *Nature* 1986;322:641–644. [PubMed: 3638500]
22. Dworetzky SI, Lanford RE, Feldherr CM. *J Cell Biol* 1988;107:1279–1287. [PubMed: 3170630]

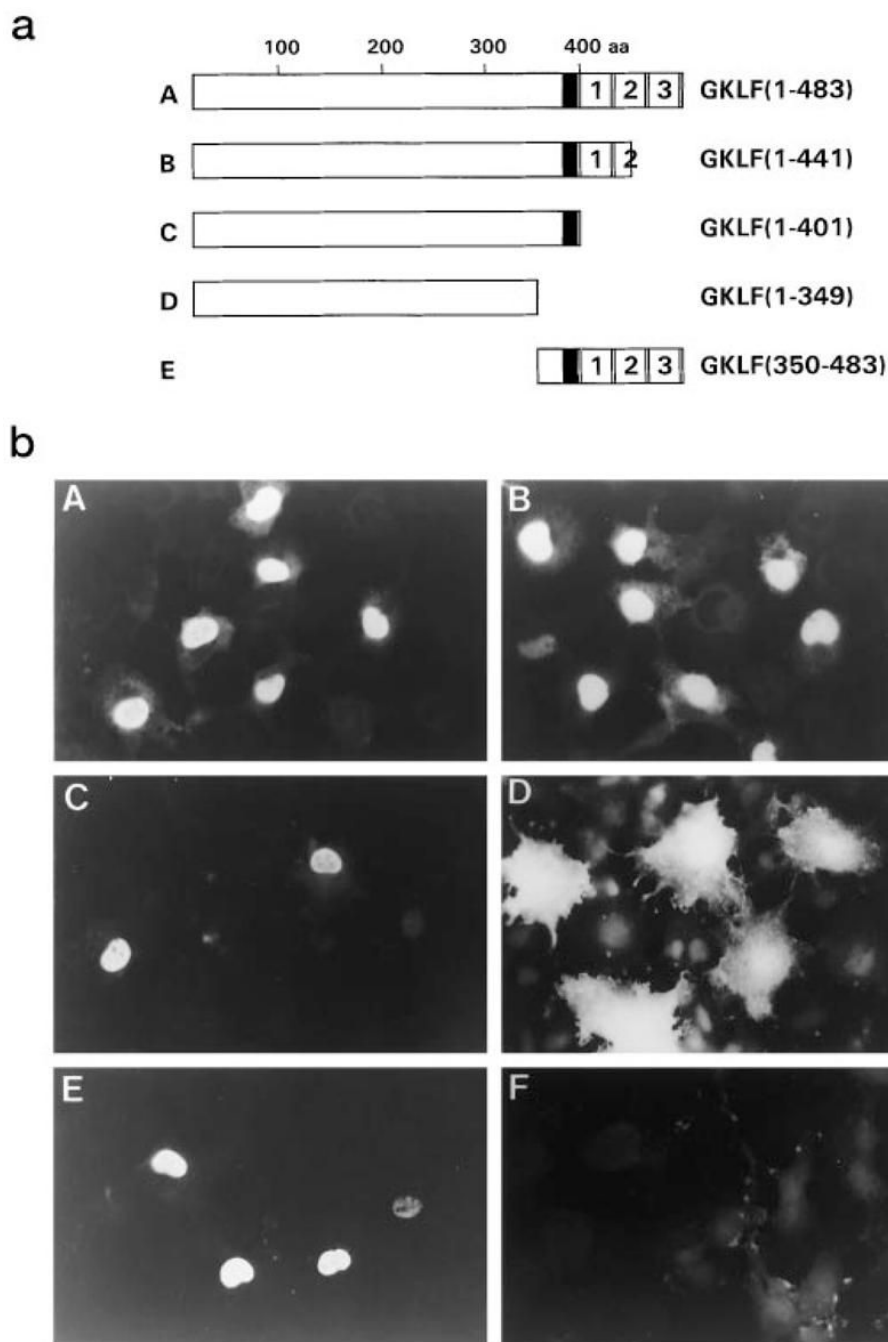


Fig. 1. Immunolocalization of GKLf in transfected COS-1 cells

a depicts the various constructs used in the experiment. The full-length GKLf is 483 aa. The 5' basic region is shown as a *filled box*. The three zinc fingers are sequentially numbered. All constructs were generated in the expression vector, PMT3. *b* is the result of indirect immunofluorescence analysis of COS-1 cells transfected with the various constructs shown in *a*. *F* is a negative control showing cells transfected with an empty PMT3 vector.

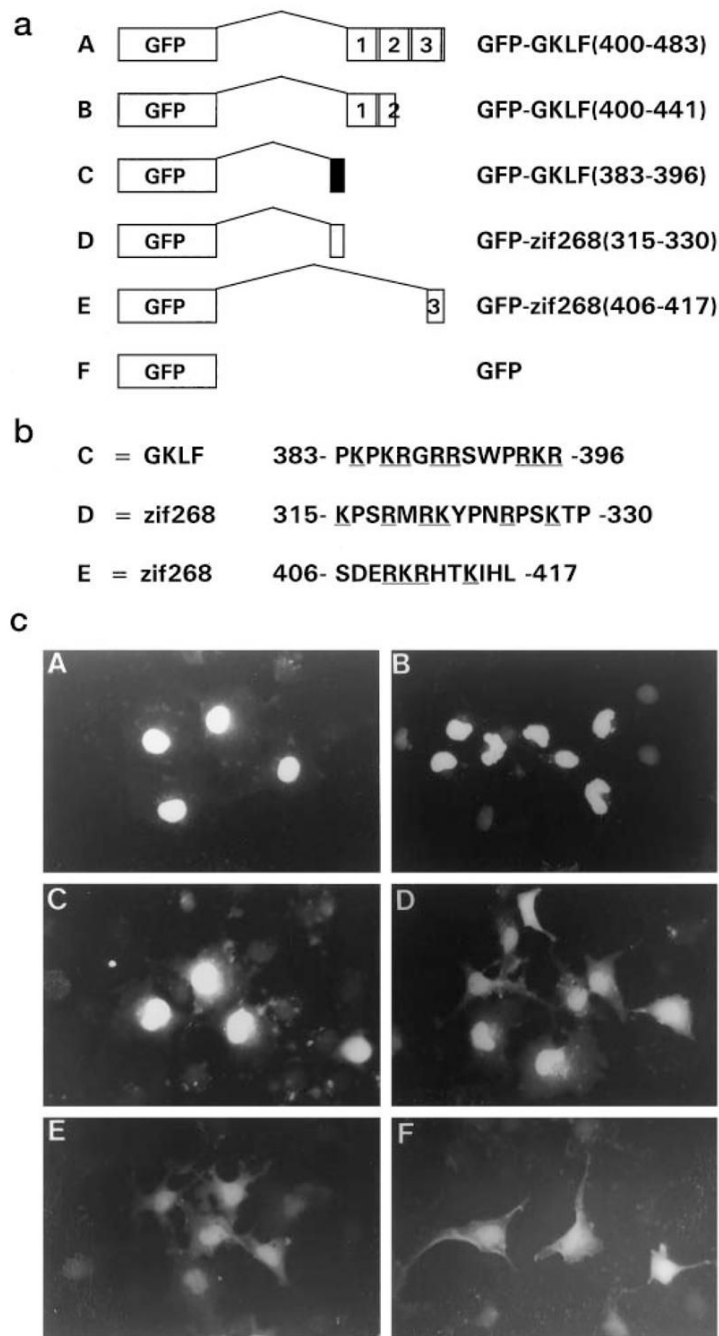


Fig. 2. Cellular localization of green fluorescent protein (GFP) fusion proteins
a shows the various GFP fusion constructs cloned in the expression vector, pEGFP-C3. The 5' basic region of GKLF is shown as a *filled box* and that of zif268/Egr-1 is shown as a *shaded box*. *Construct E* is GFP fused to a peptide containing a core NLS within the third zinc finger of zif268/Egr-1. *b* shows the amino acid sequences of peptides included in *constructs C, D, and E*. Basic amino acid residues are *underlined*. *c* shows the results of fluorescence microscopy of COS-1 cells transfected with the depicted constructs.

Name			*	Reference
mGKLF	382	E P K P K R G R R S W P R K R T A T H T C		5,6
mLKLF	253	K P K P K R G R R S W P R K R A A T H T C		7
mEKLF	261	T A P S K R G R R S W A R K R Q A A H T C		8
hBTEB2	117	Q P V R Y N R R S N P D L E K R R I H Y C		9
mBKLF	241	G K R P L P V E S P D T Q R K R R I H R C		12
hCPBP	189	K P G D K G N G D A S P D G R R R V H R C		13
hSP1	519	Y C K D S E G R A S G D P G K K K Q H I C		14
hEGR α	340	S P S A A K V T P Q I D S S R I R S H I C		15
hTIEG	351	S P S A A R V T P Q I D S S R I R S H I C		16
mGCBP	351	S P S A A R V T P Q I D S S R V R S H I C		—
mEGR-1	315	K P S R M R K Y P N R P S K T P P H E R P Y A C		17,18
mEGR-2	258	P I L R P R K Y P N R P S K T P V H E R P Y P C		19
mEGR-3	254	K P I R P R K Y P N R P S K T P K H E R P H A C		20

Fig. 3. Amino acid sequence alignment of the 5' basic regions of Krüppel-like proteins

The position of the first amino acid residue in the 5' basic region is shown for each protein. The basic amino acid residues are *boxed*. The first cysteine residue of the first zinc finger of each protein is identified by an *asterisk*. The *h* and *m* preceding the name of each protein indicates human and mouse, respectively. The study describing mouse GC box-binding protein (mGCBP) has not been published, although its sequence has been deposited in the GenBank™ data base under the accession number Z36270.