Vol. 194, No. 1, 1993 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
July 15, 1993 Pages 425-431

NGFIA (EGR1) contains transcription activating domains in both the amino terminal and carboxyl terminal regions of the protein

Christopher DeFranco[1,3], Marguerite Ro[2], Martha Grossel[2], Milton A. English[5], Ulla M. Hansen[2], John A. Wagner[1,4] and Jonathan D. Licht[2,5,*]

[1]Division of Pediatric Oncology, [2] Laboratory of Eukaryotic Transcription, Dana-Farber Cancer Institute, Harvard Medical School
Boston, MA 02115

Received June 1, 1993

We constructed and tested a number of *lac* repressor fusion proteins containing various portions of the zinc-finger conatining protein NGFIA for their ability to stimulate transcription of a reporter gene containing *lac* operators. NGFIA contains two transcription activation regions, found in two distinct regions of the protein. The carboxyl (C) terminal portion of the molecule contains a weak activation domain, including five tandem copies of an eight amino acid repeat (T/S,T/S,F/Y,P,S,P,X,X). These five tandem copies of the repeated sequence activated reporter gene transcription 4-7 fold. Amino acids 1 though 293 in the amino (N) terminus of NGFIA function as a strong transcription activation domain stimulating transcription up to 80-fold. Fusions including amino acids 1-393 failed to activate transcription,indicating the presence of a domain capable of supressing the N-terminal transcriptional activation function.

NGFIA, a rat gene also known as zif268, EGR1, Krox24, TIS8 and d2 (1-3), is an immediate early gene, transcriptionally induced by a number of agents including growth factors, cellular depolarization and activators of protein kinase A and protein kinase C. The gene product of NGFIA/EGR1 is a sequence specific DNA binding phosphoprotein with three zinc finger motifs. One distinctive features of NGFIA includes a stretch of nine serine residues followed by seven glycine residues in the amino terminal portion of the molecule. The protein also contains an eight amino acid repeat (T/S,T/S,F/Y,P,S,P,X,X) that is present three times in the amino terminus, and eight times in the carboxyl terminus (4).

Krox 24, the mouse homolog of NGFIA, activates transcription when bound to the sequence CGCCCCCGC (3,5). The activation region of NGFIA/EGR-1 has not been well defined although, transcriptional activation function was conferred to the Wilms' tumor associated

[3]Current address: Department of Neurobiology, Harvard Medical School, Boston, MA.
[4]Current address: Department of Cell Biology Cornell University Medical College, New York, NY.
[5]Current address: Brookdale Center for Molecular Biology, Mount Sinai School of Medicine, NY, NY.
*To whom correspondence should be addressed.

0006-291X/93 $4.00

protein encoded by the *WT-1* gene by formation of a fusion of the N-terminal region of EGR-1 to the zinc finger domains of *WT-1* (6) The present study further localized the activation regions of NGF1A through the formation of *lac* -repressor fusion proteins and transfer of the activation function of NGFIA to this heterologous DNA binding protein.

MATERIALS AND METHODS

Transfections- Reporter plasmid (7.5 µg), effector plasmid (7.5 µg) and an internal control pMTGH (1 µg) (7) were co-transfected along with 5µg of pBluescript plasmid (Stratagene) by the calcium phosphate precipitation method into CV-1 cells as described (8). At 48 hours post-transfection cells were harvested and assayed for CAT activity and cell media assayed for growth hormone (GH) (8). CAT activity was quantified by scanning thin layer chromatogram plates with a Betascope (Betagen) scanner. Activation is defined as the ratio of CAT activity directed by Lac/NGFIA fusion proteins relative to that observed by transfection of a plasmid expressing *lac* repressor fused to 25 amino acids of ß-galactosidase (Lac I/Z). CAT activities were normalized to the amount of growth hormone in the media to correct for transfection efficiency. Experiments were performed at least four times.

Plasmids- A plasmid (pSJT119) (9) (Gift of Steven Triezenberg) containing sequences encoding the C-terminal 80 amino acids of the acidic activator VP16 linked to herpes simplex virus thymidine kinase (HSV tk) polyadenylation sequences was digested with *Hind*III, filled in with Klenow fragment of DNA polymerase I and subsequently digested with *Bgl*II. The resulting fragment was ligated to a plasmid Lac/Kruppel 213-466 (8) which had been digested with *EcoR*I, filled in with Klenow fragment and subsequently digested with *Bam*HI to delete *Drosophila Krüppel* sequences. The resulting Lac/VP16 fusion gene served as a positive control for *trans*-activation of transcription. The LacI/Z negative control expression plasmid was constructed from Lac-VP16 by excision of VP16 sequences by restriction with *Bam*HI and *Xho*I and insertion of a synthetic duplex oligonucleotide containing stop codons in all three reading frames. Full length NGFIA cDNA was isolated from a NGF induced rat PC12 cell cDNA library (2). Lac-NGFIA plasmids were constructed by isolating DNA fragments encoding segments of the NGFIA protein and ligating them into *Xho*I /*Bam*HI digested Lac-VP16 using synthetic adaptors when necessary. The DNA fragments, their restriction sites and sequence numbers, and the amino acids encoded by these NGFIA coding fragments were as follows (4): *Nco*I(351)- *Bgl*II (1980) encoded amino acids 1-508; *Nco*I(351)-*Sph*I(1230) encoded amino acids 1-293; *Afl*II (1525)-*Bgl*II (1980) encoded amino acids 393-508 and *Nco*I(351)-*Afl*II(1525) encoded amino acids 1-393. Truncation of the N-terminal region of the NGFIA protein in Lac-NGFIA constructs was accomplished by restriction digestion of Lac-NGFIA(1-508) at various sites within the NGFIA coding region and by digestion within the synthetic linker at its 3' terminus of Lac-NGFIA(1-508). The resulting DNA fragments were ligated and recircularlized and stop codons were provided by the 3' untranlsated sequences of the HSV *tk*. gene. The DNA fragments, their restriction sites and sequence numbers, and the amino acids encoded by these NGFIA coding fragments were as follows: *Nco*I(351)-*Eco*47III(518) encoded amino acids 1-54 and *Nco*I(351)-*Dra*III(709) encoded amino acids 1-120. The Lac-NGFIA plasmid that encodes amino acids 426-473 was constructed by PCR amplification of NGFIA cDNA from nt 1629-1770 with primers containing sequences to assure the proper reading frame of the Lac/NGFIA fusion gene. The resulting PCR fragment was digested with *Xho*I and *Bam*HI and ligated to *Xho*I, *Bam*HI digested Lac-VP16. The accuracy of all constructs was verified by restriction mapping and dideoxynucleotide sequencing of double stranded plasmids.

Electrophorectic Mobility Shift Assay- CV-1 cells were transfected with 10µg of each Lac/NGFIA expression plasmids as noted. At 48 hours post-transfection, media was harvested from the cells for growth hormone as above. The cells were washed at 4°C with 140 mM NaCl, 5 mM KCl, 25 mM Tris pH 7.4 containing protease inhibitors (Leupeptin 2µg/ml, Antipain 1ug/ml, Aprotinin 5 µg/ml, Bacitracin 1µg/ml, Soybean trypsin inhibitor 50µg/ml, 1 mM Benzamidine) and scraped from the dishes. The cells were pelleted by centrifugation in a 15 ml conical tube in a Beckman JS-5.2 rotor at 1500 rpm for 10 minutes at 4°C, resuspended in 150µl of lysis buffer (50 mM Tris pH 8.0, 500 mM NaCl, 2 %v/v NP-40 + protease inhibitors as above) and incubated on ice for 30 min. The suspension was then centrifuged in a microfuge at 10,000 x g, 4°C for 20 minutes and 100µl of the supernatant was collected, avoiding the nucleic acid pellet. Protein

concentrations of the extracts were determined using the Biorad DC protein assay kit with bovine gamma globulin as the standard. The extracts ranged in concentration from 0.4 to 1.4 mg/ml, and were quick frozen in liquid nitrogen and stored at -80°C prior to use. A duplex oligonucleotide representing a high affinity symmetric *lac* operator (10) with *Sal*I overhanging ends was labeled by filling in the ends with (α-^{32}P)TTP (3000Ci/mM) and the large fragment of DNA polymerase I (New England Biolabs) for 30 minutes at room temperature followed by a 30 minute chase with a mixture of all four unlabeled nucleotides at a final concentration of 200μM. The radiolabeled duplex oligonucleotide was purified through a 6% non-denaturing polyacrylamide gel, eluted in 0.5M ammonium acetate, 1mM EDTA and precipitated by addition of $MgCl_2$ to a concentration of 10mM and addition of 3 volumes of 100% ethanol. In a 20μl volume, 2.5μg of extract from each plate of transfected cells was preincubated with 0.5μg of poly d(I-C) carrier nucleic acid, in 20mM Tris pH 7.5, 6% glycerol. 100mM NaCl for 15 minutes at 30°C. Aproximately 28 fmoles (100,000 cpm) of labeled *lac* operator probe was added to the mixture and allowed to incubate for 10 minutes at room temperature. The DNA-protein complexes were resolved on a 19:1 acrylamide:bisacrylamide, 6% non-denaturing containing polyacrylamide gel in 44.5mM Tris-borate, 44.5mM Boric acid, 1mM EDTA and run at 10V/cm for two hours. The gels were dried and autoradiographed at -80°C on Kodax XAR or Dupont Cronex film.

RESULTS AND DISCUSSION

To define the activating regions of the NGFIA protein, a number of fusion genes were constructed, containing portions of the NGFIA cDNA and the cDNA of a heterologous DNA binding protein, the *lac* repressor. A similar fusion between the potent transcription activator VP16 (11) and the *lac* repressor was used as a positive control. To assay the fusion proteins for transcriptional activity, plasmids expressing each Lac/NGFIA fusion gene were co-transfected into CV-1 cells along with a plasmid in which five *lac* operators were inserted upstream of the herpes simplex virus thymidine kinase promoter linked 5' to a chloramphenicol acetyltransferease (CAT) reporter gene (8). To determine whether the expression vectors yielded Lac/NGFIA protein when transiently expressed, electrophoretic mobility shifts assays (EMSA) were performed with extracts prepared from cells transfected with each of these expression vectors. Briefly, this assay measured protein expression by the ability of the Lac/NGFIA fusion proteins to retard the mobility of a DNA fragment containing a *lac* operator. A Lac/NGFIA fusion protein containing amino acids 1-293 of NGFIA caused a 10-80 fold increase in the CAT activity relative to constructs containing no NGFIA sequence (Fig. 1, 2 and 4), while a Lac/NGFIA fusion containing amino acids 1-393, including the zinc finger region of NGFIA did not activate transcription (Fig. 1). In comparison, a Lac/VP16 fusion protein stimulated transcription from 23-59 fold under identical conditions (Data not shown). Both Lac/NGFIA1-393 and Lac/NGFIA 1-293 yielded weak but detectable, novel DNA-protein complexes with a radiolabeled *lac* operator DNA fragment, compared to control extracts derived from cells transfected with only pBluescript (Stratagene) plasmid DNA (Fig. 3). The Lac/NGFIA 1-293 extract yielded three complexes, one with the same mobility as the Lac/NGFIA 1-393 complex and two others with slower electrophorectic mobility. This suggests the formation of a higher order complex, perhaps indicating the formation of dimers or tetamers of the Lac/NGFIA 1-293 protein. The inability of Lac/NGFIA 1-393 to activate transcription suggests that sequences between amino acids 293 and 393, which include the zinc finger region of the NGFIA protein (found between amino acids 313 and 393) may act to dampen or supress the transcriptional activation region within the N-terminal portion of the NGFIA protein. To further

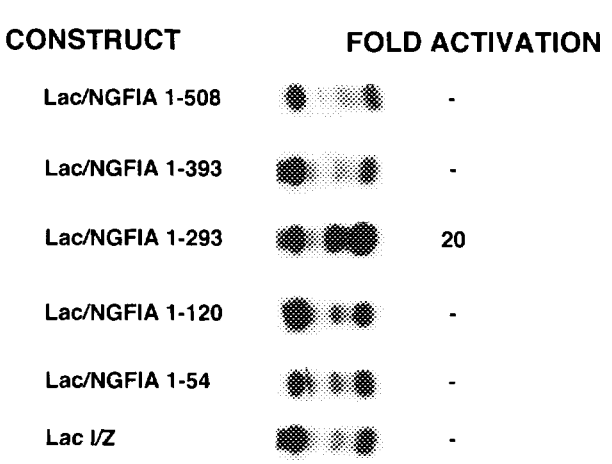

Figure 1. Transcriptional activation by the N-terminal domain of the NGFIA protein.
CV-1 cells were transfected with a reoprter gene containing five *lac* operators upstream of the herpes simplex virus thymidine kinase promoter, along with expression vectors for Lac/NGFIA fusion genes as indicated. Fold activation relative to co-transfection of the reporter plasmid with a plasmid expressing the *lac* repressor itself is indicated to the right of the chromatogram.

map the amino (N)-terminal activation region, two additional fusion genes were tested. Lac/NGFIA 1-120 and 1-54 both failed to activate transcription. By EMSA, very large quantities of Lac/NGFIA 1-54 DNA were found in transfected CV-1 cells but no detectable DNA binding by Lac/NGFIA 1-120 was noted (Fig. 3). Assay of GH from the media of transfected cells indicated both transfections were of similar efficiency. Together these data indicate the presence of a strong transcriptional activation region between amino acids 54 and 293 of NGFIA and a potential negative regulator of this activation domain located between amino acids 293 and 393.

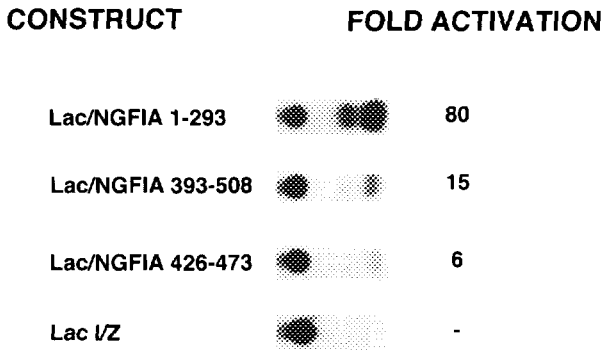

Figure 2. Transcriptional activation by the C-terminal domain of the NGFIA protein.
CV-1 cells were transfected with a reoprter gene containing five *lac* operators upstream of the herpes simplex virus thymidine kinase promoter, along with expression vectors for carboxyl terminal Lac/NGFIA fusion genes. Fold activation relative to co-transfection of the reporter plasmid with a plasmid expressing the *lac* repressor itself is indicated to the right of the chromatogram.

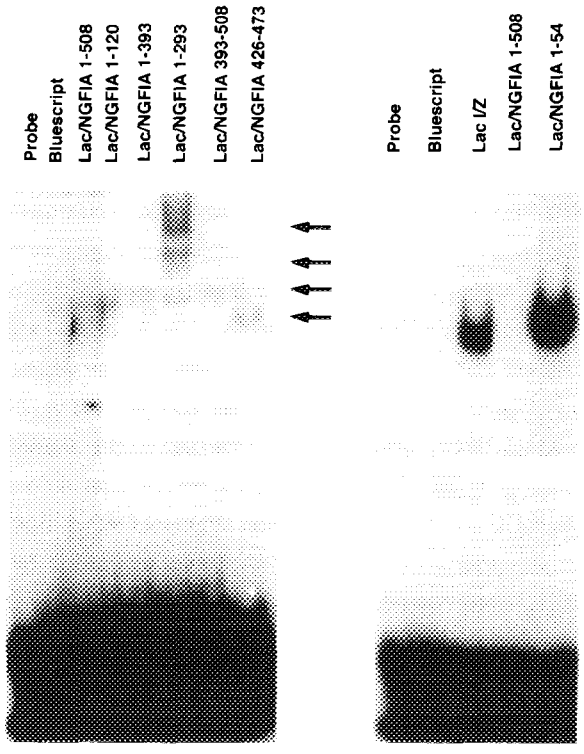

Figure 3. Electrophoretic mobility shift assay with extracts from cells transfected with Lac/NGFIA fusion genes.

CV-1 cells were transfected with 10μg of each Lac/NGFIA expression vector and whole cell extracts were prepared as described in materials and methods. A radiolabeled duplex oligonucleotide containing the *lac* operator sequence was incubated with 2.5μg of each extract and DNA-protein complexes were separated by non-denaturing gel electrophoresis. Arrows indicate novel DNA-protein complexes. The panel on the left was exposed to 13 times more ^{32}P decay than the right panel.

A full length fusion between the *lac* repressor and the NGFIA gene did not lead to transcriptional activation, despite the observation by others that the NGFIA protein activates transcription through a GC rich site (3,5). However, EMSA indicated that this construct did not direct the production of detectable amounts of Lac fusion protein capable of binding to the *lac* operator *in vitro*. This could be due to the instability of the resulting fusion protein. Alternatively, the fusion protein may be incapable of binding to the *lac* operator. A precident for this possibility exists as one group found that fusion of certain protein segments to the DNA binding domain of the yeast *GAL4* protein blocked the ability of the *GAL4* moeity to bind to its target sequence (12).

The carboxyl (C)-terminus of NGFIA was also examined for the presence of transcription activation regions. Lac/NGFIA 393-508 and Lac/NGFIA 426-473 fusion proteins stimulated transcription 15 and 6 fold repectively in this particular experiment, while the Lac/NGFIA 1-293 stimulated transcription 80-fold from the *lac* operator containing reporter (Fig. 2). This indicated the presence of a significant, but much weaker activation domain in the C-terminal portion of the NGFIA protein. EMSA assay with extracts prepared from cells transfected with both of these

fusion genes indicated the presence of *lac* operator-protein complexes, consistant with the production of similar amounts of fusion protein (Fig. 3).

The amino acid sequence of the N-terminus of NGFIA, a region that contains a strong transcription stimulating activity, does not reveal obvious similarity to known acidic, glutamine rich or proline rich activation domains (13) but is serine-rich like the transcriptional activation region of the Pit-1 POU-homeodomain containing transcription factor (14). A common feature of both the N-terminal and C-terminal transcription activation domains found in NGFIA is the conserved eight amino acid repeat. Lac/NGFIA 426-473 contains five copies of the conserved eight amino acid repeat sequence and Lac/NGFIA 393-508, a somewhat stronger transcriptional activator than Lac/NGFIA 426-473, includes seven copies of the repeat. Both of these fusion proteins produced only an average 4-7 fold transcription activation (Fig. 2 and 4). The N-terminal activation region contains three copies of the conserved eight amino acid repeat sequence. Therefore, the strong activation (30-80 fold) directed by the N-terminal domain cannot be attributed to the conserved, repeated sequences. Other protein segments either alone or in concert with the repeat sequence must be needed for strong transcriptional activation by NGFIA. Possibly the serine and threonine rich repeats could serve to modulate the activity of the activation domain(s)

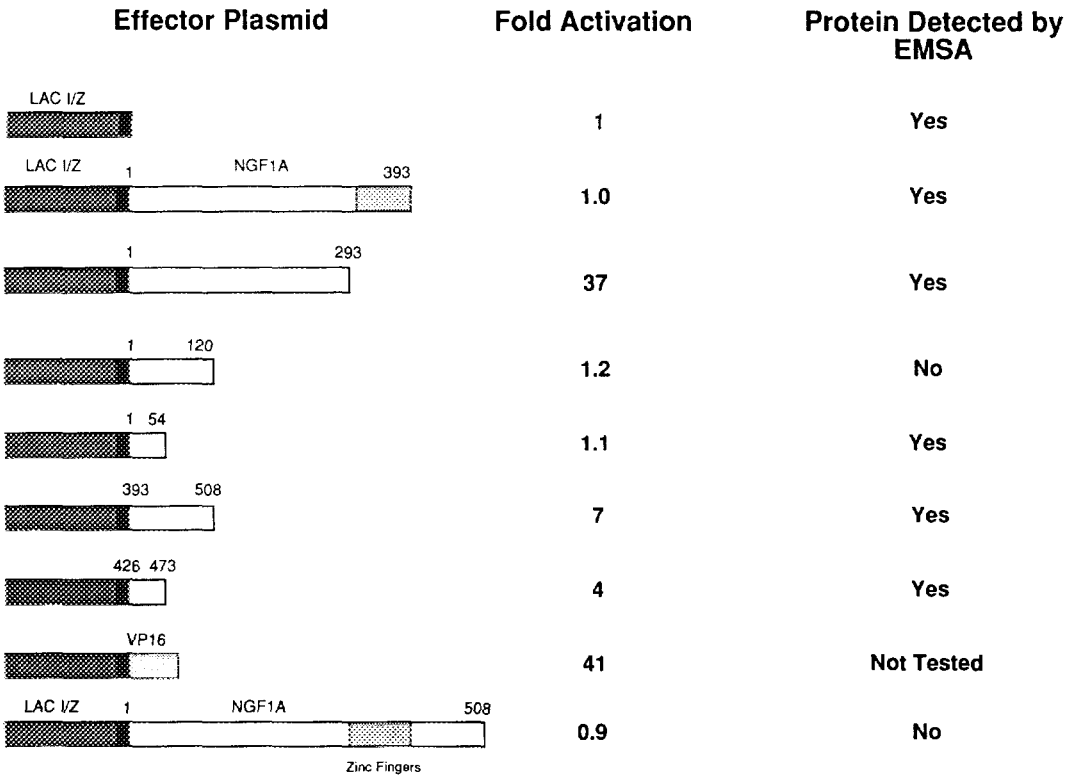

Figure 4. Summary of the structure of Lac/NGFIA fusion proteins, their effect on transcription and their ability to be detected in transfected cells by EMSA. The fold activation is the average of 4 to 8 independent experiments.

through phosphorylation, as in the case of the cyclic AMP responsive factor CREB (15). However, phosphorylation sites within NGFIA protein have yet to be defined.

Like other transcriptional activators such as C/EBP (16), there also appears to be a domain within NGFIA that represses the ability of the activation domain to function. This domain may be latent or regulated in activity in the context of the wild-type protein and has become unmasked in the context of the *lac* repressor fusion protein. There is no evident amino acid similarity between this negative domain and found in C/EBP.

NGFIA appears to be an example of a protein where the transcription activating regions are distributed over separable regions of the protein (13). It is possible that the activation region in the caboxyl terminus could cooperate with the strong activation region in the amino terminal portion of the protein to yield full activation of transcription by the native NGFIA protein.

ACKNOWLEDGMENTS

Supported by Physician Scientist Award CA 01272 and a Biomedical Research Support Grant (JL); NRSA Fellowship CA 0880 (CD) and CA22427 (JAW). This is publication #133 from the Brookdale Center for Molecular Biology.

REFERENCES

1. Hershman, H. R. (1991) Annu. Rev. Biochem., 60, 281-319
2. Cho, K. C., Scarnes, W. S., B. Minsk, B., Palmieri,S, Jackson-Grusby, L., and Wagner, J. A. (1989) Mol. Cell Biol. 9, 135-143
3. Sukhatme V. P. Am J Kidney Dis 17: 615-618, 1991
4. Milbrandt, J. (1987) Science 238, 797-799
5. Lemaire, P., Vesque,C. , Schmitt, J., Stunnenberg, H., Frank, R., and Charnay,P. (1990) Mol. Cell. Biol. 10, 3456-3467
6. Madden, S. L., Cook, D.M., Morris, J.F., Gashler, A., Sukhatme, V.P., and Rauscher, F.J. III. (1991) Science 253: 1550-1553
7. Selden, R. F., Burke-Howie, K., Rowe, M. E., Goodman, H. M., and Moore, D. D. (1986) Mol. Cell Biol. 6, 3173-3179
8. Licht, J. D., Grossel, M. J., Figge, J., and Hansen, U. M. (1990) Nature 346, 76-79
9. Triezenberg, S.J., Kingsbury,. R. C., and McKnight, S.L. (1988) Genes & Dev. 2, 718-229
10. Figge, J. M., Wright, C., Collins, C. J., Roberts, T. M., and Livingston, D. M. (1987) Cell 52:713-722
11. Sadowski, I., Ma, J., Triezenberg, S., and Ptashne, M. (1988) Nature 335: 563-564
12. Golemis, E. A., and Brent, R. (1992) Mol Cell Biol 12:3006-3014
13. Mitchell, P. J., and Tjian, R. (1989) Science. 245, 371-378
14. Theill, L. E., Castrillo, J-L., Wu, D., and Karin, M. (1989) Nature 342: 945-948
15. Montminy, M. R., Gonzales, G. A., and Yamamoto, K. K. (1990) Trends Neurosci. 13, 184-188
16. Pei, D., and Shih, C. (1991) Mol. Cell Biol. 11, 1480-1487