aiding in its installation, M. Levitt for numerous helpful and critical discussions, Z. Wasserman for important contributions to code development and methodology, and R. Hilmer for the interactive computer graphics used in the analysis of molecular dynamics simulations.

29 June 1987; accepted 1 September 1987

A Nerve Growth Factor–Induced Gene Encodes a Possible Transcriptional Regulatory Factor

JEFFREY MILBRANDT

Nerve growth factor (NGF) is a trophic agent that promotes the outgrowth of nerve fibers from sympathetic and sensory ganglia. The neuronal differentiation stimulated by this hormone was examined in the NGF-responsive cell line PC12. Differential hybridization was used to screen a complementary DNA library constructed from PC12 cells treated with NGF and cycloheximide. One of the complementary DNA clones that was rapidly induced by NGF was found to have a nucleotide sequence that predicts a 54-kilodalton protein with homology to transcriptional regulatory proteins. This clone, NGFI-A, contains three tandemly repeated copies of the 28– to 30–amino acid "zinc finger" domain present in *Xenopus laevis* TFIIIA and other DNA-binding proteins. It also contains another highly conserved unit of eight amino acids that is repeated at least 11 times. The NGFI-A gene is expressed at relatively high levels in the brain, lung, and superior cervical ganglion of the adult rat.

NERVE GROWTH FACTOR (NGF) IS a polypeptide hormone that is required for the development and survival of sympathetic and neural crest–derived sensory neurons in vivo and in vitro (*1*). NGF also participates in nerve regeneration after injury (*2*) and plays a role in the central nervous system (*3*). The effects of NGF on the differentiation of sympathetic neurons are manifested by an accelerated outgrowth of nerve fibers and the induction of several enzymes involved in neurotransmitter biosynthesis (*4*). During the differentiation process, specific genes are turned on and off at precise times. This ordered pattern of gene expression, resulting in the establishment of the neuronal phenotype, is presumably controlled by DNA-binding proteins that act to regulate transcription.

The best characterized transcriptional activating factor is the TFIIIA protein that regulates the expression of 5*S* RNA genes during development in *Xenopus laevis* (*5*). The DNA-binding region of this protein is composed of nine tandemly repeated units that contain two invariant pairs of Cys and His residues (*6*). Each repeat unit is folded around a single zinc ion that is coordinately bonded to the Cys and His residues to form a DNA-binding domain or "Zn finger." Amino acids capable of interacting with the DNA helix (Lys, His, Asn, Gln, Thr, and Arg) are located at the tip of each "Zn finger." This 28– to 30–amino acid DNA-binding domain has now been identified in several other proteins involved in transcriptional regulation, including yeast ADR1-encoded protein (*7*) and the Kruppel gene of *Drosophila* (*8*).

One model of neuronal differentiation is the PC12 cell line (*9*), whose NGF-mediated transition from replicating adrenal chromaffin-like cells to sympathetic neuron-like cells is prevented by the addition of RNA synthesis inhibitors (*10*). This cell line, derived from a rat pheochromocytoma, responds to NGF by extending neurites and by increasing the transcription of several genes, including ornithine decarboxylase (*11*), GAP43 (*12*), and c-*fos* (*13*). The activation of the proto-oncogene c-*fos* is rapid but transient in NGF-treated PC12 cells. The c-*fos* transcripts accumulate to even higher levels in cells stimulated in the presence of cycloheximide (CHX), in part because of an increase in messenger RNA (mRNA) half-life (*14*). Using c-*fos* as a model of an early activated gene, I have taken advantage of this CHX-mediated superinduction to identify other early, NGF-induced genes that may regulate neuronal differentiation.

A complementary DNA (cDNA) library containing 8×10^5 independent clones was constructed in the vector λgt10 (*15*), by using mRNA isolated from PC12 cells that had been treated with NGF (50 ng/ml) and CHX (10 μg/ml) for 3 hours. Initially, 8000 recombinants (1% of total library) were screened by differential hybridization with single-stranded cDNA probes complementary to mRNA from either unstimulated PC12 cells (no NGF) or PC12 cells treated with NGF and CHX for 3 hours. Three different NGF-induced cDNAs were identified from this screen: c-*fos*, a protein that binds DNA (*16*) and may function as a trans-acting factor (*17*), NGFI-A, and NGFI-B.

NGFI-A was selected for further study. This cDNA hybridized to a ~3.3-kb mRNA that was induced by NGF (Fig. 1A, lane 3) but accumulated to a much higher level in PC12 cells treated with NGF and CHX (lane 4). In PC12 cells grown without NGF this mRNA could not be detected (lane 1). However, when cells were treated with CHX alone, the level of this transcript was slightly increased (lane 2).

The time course of induction of the NGFI-A mRNA was determined by isolating mRNA from PC12 cells treated with NGF for various lengths of time. Northern analysis with the NGFI-A cDNA probe showed that the basal level of this transcript was very low in PC12 cells (Fig. 1B), but within 15 minutes after the addition of NGF (lane 2) the level of NGFI-A mRNA

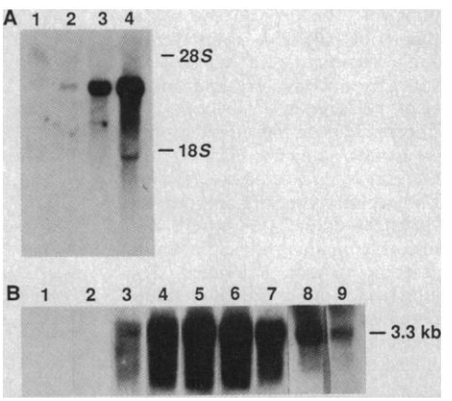

Fig. 1. Identification of an NGF-regulated cDNA clone. Total cellular RNA was isolated by the method of Chirgwin (*24*). The cDNA library was screened by differential hybridization with single-stranded cDNA probes complementary to mRNA from either naïve PC12 cells (no NGF) or PC12 cells treated with NGF and CHX for 3 hours (*25*). Phage from the NGFI-A cDNA clone were isolated on DEAE columns (*26*). DNA was purified and digested with Eco RI. The cDNA insert was isolated by electrophoresis on low-melting agarose and labeled with ^{32}P-labeled deoxyadenosine triphosphate by oligo labeling (*27*). RNA electrophoresis and Northern blot analysis were performed as described (*13*). (**A**) A Northern blot containing RNAs from PC12 cells treated for 3 hours with (lane 1) no addition, (lane 2) CHX (10 μg/ml), (lane 3) NGF (50 ng/ml), or (lane 4) NGF and CHX, was probed with the NGFI-A cDNA. The locations of the 28*S* and 18*S* ribosomal RNAs are indicated. (**B**) Time course of the NGF-mediated induction of NGFI-A mRNA. RNA samples from PC12 cells grown in the presence of NGF for 0, 2, 15, 30, 45, 60, and 120 minutes, 13 hours, and 6 days (lanes 1 through 9, respectively) were hybridized to ^{32}P-labeled NGFI-A cDNA insert. The 3.3-kb NGFI-A mRNA is indicated.

Division of Laboratory Medicine, Departments of Pathology and Internal Medicine, Washington University School of Medicine, St. Louis, MO 63110.

increased. The NGFI-A mRNA level peaked approximately 30 to 45 minutes after the addition of NGF (lanes 3 and 4) and persisted for up to 6 days (lane 9). The rate of induction of the NGFI-A gene was similar to that of the NGF induction of c-*fos* (*13*), but the rate of decay of NGFI-A was much slower. The NGFI-A gene was also induced by treating PC12 cells with the phorbol 12-myristic 13-acetate (PMA) and the calcium ionophore A23187 as has been demonstrated for the c-*fos* gene (*18*).

As shown in Fig. 2, the NGFI-A cDNA is 3118 nucleotides in length and contains an open reading frame from nucleotide 269 to nucleotide 1876. Several in-frame Met codons are present, but only the Met at nucleotide 353 contains the translational initiation consensus sequence (CCG/ACCATGG) (*19*). On the basis of initiation at this site, the NGFI-A protein is 508 amino acids long with an unmodified molecular size of 53,934 kD. Consensus polyadenylation sites are located at nucleotide 1920, 3132, and 3143. Multiple copies of the sequence ACAAAA are also present; this sequence has been postulated to provide a protein-binding domain in either the gene or the corresponding mRNA of the Kruppel gene (*8*).

A search of the NBRF protein database (*20*) revealed that the NGFI-A protein is homologous to the *X. laevis* transcriptional factor, TFIIIA. As shown in Fig. 2, the region of NGFI-A that is homologous to TFIIIA extends from nucleotide 1277 to nucleotide 1534. This region contains three 28– to 30–amino acid repeats. In the TFIIIA molecule, the area of homology contains nine copies of a similar 28– to 30–amino acid repeat. In Fig. 3A, the three NGFI-A repeat sequences are aligned with the TFIIIA repeat and similar sequences from other DNA-binding proteins that regulate transcription. The high degree of similarity is apparent from this comparison. The most highly conserved residues are boxed; these include the residues determined to be most important in the nine TFIIIA repeats.

The NGFI-A sequence contains several other unique features. It contains a stretch of predominantly basic residues which extends from Arg 380 to Lys 401. Immediately following this basic region (at nucleotide 1556 and extending to the COOH-terminus) is a hydrophilic region that is highly enriched (59 of 107 residues) with Pro, Ser, and Thr. Within this COOH-terminal portion of the molecule another tandemly repeated unit can be distinguished. This is illustrated in Fig. 4, where the predicted amino acid sequence is compared to itself and displayed in a dot matrix format (*21*). The central, continuous diagonal results from the perfect match of NGFI-A protein with itself; the short diagonals offset from the central one represent repeat units. One set of three repeat units is located between residues 309 and 394 and represents the "Zn finger" domains discussed above. Another site of repeating units can be seen between residues 413 and 487. This repeat has a periodicity of eight amino acids and is reiterated eight times. Three additional copies of this repeat are located in the NH_2-terminal region of the molecule (Fig. 2). These repeats are aligned with each other, along with a consensus sequence, in Fig. 3B.

The NH_2-terminal portion of the NGFI-A molecule is also composed of a high percentage of Pro, Ser, and Thr (45%) from nucleotide 455 to nucleotide 1045. A potential phosphorylation site (Thr/Ser-Gly-Arg) is present at nucleotide 668. Several potential *N*-linked, and many potential *O*-linked, glycosylation sites are also present. A curious feature of this region is the stretch of nine Ser residues directly followed by seven Gly residues (Fig. 2). Interestingly, the same codon, AGC, is used for each of the nine Ser residues in this stretch. This triplet repeat, called *opa*, has been previously recognized in the *Notch* gene, a gene that is involved in *Drosophila* neurogenesis (*22*). The poly(Gly) residues are encoded by a repeat cluster called *pen*, which consists of clusters of GGN (where N is any nucleotide) (*23*). The juxtaposition of these two triplet repeats may indicate that the critical functional feature of this region is the underlying nucleotide sequence, rather than the amino acid sequence. The repeating AGC or GGN motifs could represent protein binding sites in the gene or in the RNA transcript.

To examine the expression of the NGFI-A

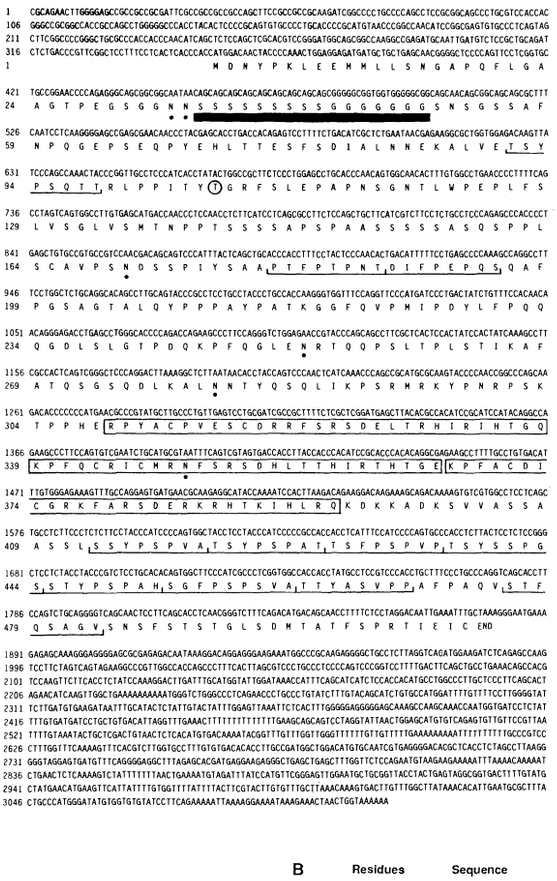

Fig. 2. Nucleotide and deduced amino acid sequence of the NGFI-A cDNA. The nucleotide sequence was determined by the method of Sanger *et al.* (*28*). The NGFI-A cDNA was cloned in both orientations into the Eco RI site of the Bluescript plasmid (Stratagene), and deletions were constructed with Exonuclease III and mung bean nuclease as described (*29*). The numbers on the left refer to the nucleotide sequence (upper) and the amino acid sequence (lower). The region corresponding to the Zn-finger domains is boxed. The regions containing the eight amino acid repeats are underlined. The long stretch of Ser and Gly residues is underscored by a heavy bar. A potential phosphorylation site is circled. Potential sites for *N*-linked glycosylation are dotted.

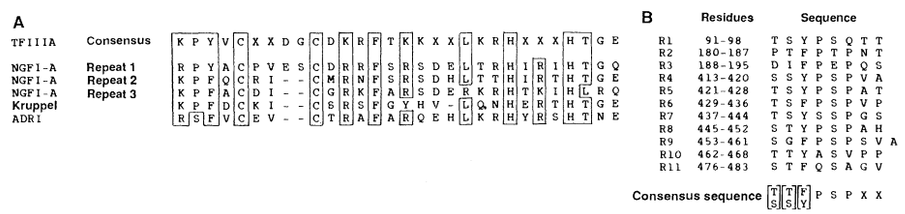

Fig. 3. (**A**) Alignment of the sequence of amino acids of the "Zn finger" repeats in the NGFI-A protein with themselves and with the DNA-binding regions of several eukaryotic transcriptional regulatory proteins. (**B**) Alignment of the highly conserved eight–amino acid repeat of the NGFI-A gene. A consensus sequence is derived for positions at which at least 8 of the 11 repeats contain identical or conservatively substituted residues.

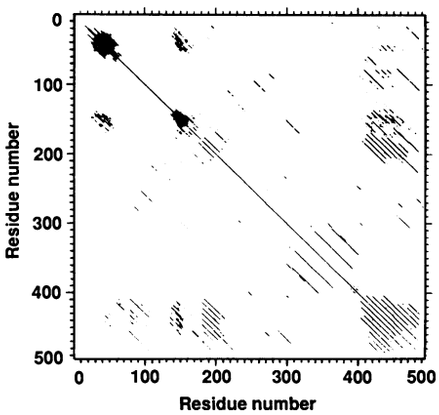

Fig. 4. Dot matrix analysis of the amino acid sequence of the NGFI-A protein. The amino acid sequence is represented on both the horizontal and vertical axes. Each 30-residue segment is compared to all other 30-residue segments. A homology score is calculated for each comparison using the mutation data scoring matrix (*20*). A dot is placed at the position given by the midpoint of the two segments when the comparison score is greater than 20.

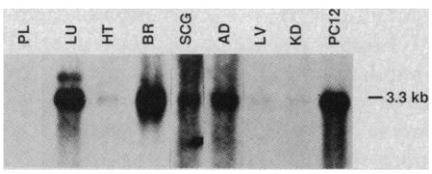

Fig. 5. Expression of the NGFI-A gene in the adult rat. Poly(A)$^+$ RNA (7.5 μg) from the indicated tissues of 6-week-old Sprague-Dawley rats was electrophoresed and transferred to nitrocellulose. The RNA blots were hybridized to the NGFI-A cDNA. In the SCG lane only 12 μg of total RNA was analyzed. The lane labeled PC12 contains 250 ng of poly(A)$^+$ RNA from PC12 cells treated with NGF and CHX. Tissues: PL, placenta; LU, lung; HT, heart; BR, brain; SCG, superior cervical ganglia; AD, adrenal; LV, liver; and KD, kidney.

gene, RNA transfer analysis was performed on samples from various tissues of the adult rat. The NGFI-A gene is expressed at high levels in the brain, lung, and superior cervical ganglia, and at a moderate level in the adrenal gland; however, the amount of NGFI-A mRNA is much greater in PC12 cells treated with NGF and CHX than in any of the tissues tested (Fig. 5). In other tissues NGFI-A transcripts are detectable but very low in number. The signal detected in the SCG lane reflects a high level of expression in these cells, since approximately one-tenth as much mRNA was loaded in this sample (12 μg of total RNA versus 7.5 μg of polyadenylated [poly(A)$^+$] RNA for the other tissues). The detection of NGFI-A transcripts in the brain is interesting in light of recent evidence that NGF-responsive neurons are present within the central nervous system (*3*). The expression in the lung and the detection of a second, slightly larger, transcript in this organ is not understood. The widespread expression observed for the NGFI-A gene opens the possibility that this gene may be regulated by factors other than NGF in some tissues.

The homology of the protein encoded by the NGFI-A gene to other transcriptional regulatory proteins suggests that it may have a similar function. The recent demonstration that c-*fos* is a DNA-binding protein, coupled with the discovery of a third NGF-induced gene, the NGFI-B gene, which is also homologous to transcriptional regulatory proteins, supports the hypothesis that these rapidly induced genes encode proteins that may play a role in initiating the differentiation program elicited by NGF. It is likely that genes that are rapidly induced in response to other environmental cues or hormones may also encode transcriptional regulatory factors.

REFERENCES AND NOTES

1. B. Yankner and E. Shooter, *Annu. Rev. Biochem.* **51**, 845 (1982).
2. S. Korsching, R. Heumann, A. Davies, H. Thoenen, *Soc. Neurosci. Abstr.* **12**, 1096 (1986); C. E. Bandtlow, R. Heumann, M. E. Schwab, H. Thoenen, *EMBO J.* 6, 891 (1987); M. Taniuchi, H. B. Clark, E. M. Johnson, Jr., *Proc. Natl. Acad. Sci. U.S.A.* **83**, 4094 (1986).
3. A. Szutowicz, W. Frazier, R. Bradshaw, *J. Biol. Chem.* **251**, 1516 (1976); T. H. Large *et al.*, *Science* **234**, 352 (1986); S. Korsching, G. Auberger, R. Heumann, J. Scott, H. Thoenen, *EMBO J.* **4**, 1389 (1985); S. Whittemore *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 817 (1986); M. Taniuchi, J. B. Schweitzer, E. M. Johnson, Jr., *ibid.*, p. 1950.
4. H. Thoenen, Y.-A. Barde, D. Edgar, in *Neurosecretion and Brain Peptides*, vol. 28 of *Advances in Biochemical Psychopharmacology*, J. B. Martin, S. Reichlin, K. L. Bick, Eds. (Raven, New York, 1981), pp. 263–272.
5. D. Engelke, S.-Y. Ng, B. S. Shastry, R. G. Roeder, *Cell* **19**, 717 (1980).
6. J. Miller, A. D. McLachlan, A. Klug, *EMBO J.* **4** 1609 (1985).
7. T. A. Hartshorne, H. Blumberg, E. T. Young, *Nature (London)* **320**, 283 (1986).
8. U. B. Rosenberg *et al.*, *ibid.* **319**, 336 (1986).
9. L. Greene and A. Tischler, in *Advances in Cellular Neurobiology*, S. Federoff and L. Hertz, Eds. (Academic Press, New York, 1982), vol. 3, pp. 373–414.
10. D. E. Burstein and L. A. Greene, *Proc. Natl. Acad. Sci. U.S.A.* **75**, 6059 (1978).
11. S. C. Feinstein, S. L. Dana, L. McConlogue, E. M. Shooter, P. Coffino, *ibid.* **82**, 5761 (1985).
12. L. R. Karns, S.-C. Ng, J. A. Freeman, M. C. Fishman, *Science* **236**, 597 (1987).
13. T. Curran and J. I. Morgan, *ibid.* **229**, 1265 (1985); M. Greenberg, L. Greene, E. Ziff, *J. Biol. Chem.* **260**, 14101 (1985); W. Kruijer, D. Schubert, I. M. Verma, *Proc. Natl. Acad. Sci. U.S.A.* **82**, 7330 (1985); J. D. Milbrandt, *ibid.* **83**, 4789 (1986).
14. K. Kelly, B. Cochran, C. Stiles, P. Leder, *Cell* **35**, 603 (1983).
15. T. V. Huynh, R. A. Young, R. W. Davis, in *DNA Cloning*, D. M. Glover, Ed. (Practical Approach Series, IRL Press, Oxford, England, 1985), vol. 1, pp. 49–78. The cDNA library was constructed as follows: poly(A)$^+$ RNA was selected by two passages over oligo(dT)-cellulose. Single-stranded cDNA was synthesized from 5 μg of poly(A)$^+$ RNA with 40 units of reverse transcriptase (Life Sciences) in the presence of 100 m*M* Tris, *p*H 8.3 at 25°C, 140 m*M* potassium acetate, 10 m*M* magnesium acetate, oligo(dT) (80 μg/ml), 800 μ*M* deoxynucleoside 5′-triphosphates by incubation at 37°C for 1 hour. The second strand was synthesized according to the method of H. Okayama and P. Berg [*Mol. Cell. Biol.* **2**, 161 (1982)]. The double-stranded cDNAs were treated with Eco RI methylase, treated with T4 polymerase, ligated to Eco RI linkers, digested with Eco RI, and size-selected by chromatography on BioGel A5M. The size-selected cDNAs were ligated to Eco RI–digested λgt10 and packaged in vitro (Stratagene). The resulting phage were used to infect *Escherichia coli* C600 Δ hfl.
16. L. C. Sambucetti and T. Curran, *Science* **234**, 1417 (1986).
17. C. Setoyama, R. Frunzio, G. Liau, M. Mudryj, B. de Crombrugghe, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 3213 (1986).
18. J. Milbrandt, unpublished data.
19. M. Kozak, *Cell* **44**, 283 (1986).
20. *Protein Sequence Data Base of the Protein Identification Resource* (National Biomedical Research Foundation, no. 10.0, Washington, DC, 1986).
21. M. O. Dayhoff, W. C. Barker, L. T. Hunt, *Methods Enzymol.* **91**, 524 (1983).
22. K. A. Wharton, B. Yedvobnick, V. G. Finnerty, S. Artavanis-Tsakonas, *Cell* **40**, 55 (1985).
23. S. R. Haynes *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 1819 (1987).
24. J. Chirgwin, A. Przybyla, R. MacDonald, W. Rutter, *Biochemistry* **18**, 5294 (1979).
25. The cDNA library was screened by differential hybridization as follows: the phage were plated at low density (1600 plaque-forming units per 88 mm^2), and four replica filters were prepared from each plate. Probes were prepared by synthesizing single-stranded cDNA from 2 μg of poly(A)$^+$ RNA from either naive PC12 cells (no NGF) or PC12 cells treated with NGF and CHX. Synthesis conditions for cDNA were similar to above with the exception that 100 μCi of ^{32}P-labeled deoxycytidine 5′-triphosphate were included. The RNA template was removed by incubation at 68°C for 1 hour with 50 m*M* NaOH, and unincorporated nucleotides were removed by centrifugation in a Centricon-30 (Amicon). Two filters from each plate were hybridized with 1 × 10^6 cpm/ml of either type of single-stranded cDNA probe in 5× standard saline citrate, 50% formamide, 5× Denhardt's, single-stranded DNA (100 μg/ml), 0.1% SDS, polyadenylic acid (1 μg/ml), at 42°C for 48 hours. Filters were washed twice for 10 minutes in 2× standard saline citrate and 0.1% SDS at 25°C, followed by two 60-minute washes in 0.2× standard saline citrate and 0.1% SDS at 65°C. Plaques that produced a differential signal were identified by inspection and picked for secondary screening.
26. G. Helms, M. Graham, J. Dutchik, M. Olson, *DNA (New York)* **4**, 39 (1985).
27. A. Feinberg and B. Vogelstein, *Anal. Biochem.* 67, 15 (1984).
28. F. Sanger, S. Nicklen, A. R. Coulson, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5463 (1977).
29. S. Henikoff, *Gene (Amsterdam)* **28**, 351 (1984).

30. I thank T. Fahrner for technical assistance and P. Manning, E. Johnson, and M. Thomas for advice. This work was supported by grant RG1779-A from the Multiple Sclerosis Foundation and grant NS01018 from the National Institute of Neurological and Communicative Diseases and Stroke.

4 June 1987; accepted 28 August 1987