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.

erve growth factor (NGF) is a polypeptide hormone that is required for the development and survival of sympathetic and neural crestderived 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 5S 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 DNAbinding domain has now been identified in

several other proteins involved in transcriptional regulation, including yeast ADR1encoded 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, NGFinduced genes that may regulate neuronal differentiation.

A complementary DNA (cDNA) library containing 8×10^5 independent clones was constructed in the vector $\lambda gt10$ (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 \sim 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 singlestranded cDNA probes complementary to mRŇA 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 ³²P-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 28S and 18S 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 ³²P-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 NGFI 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 12myristic 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/ACCAT-GG) (19). On the basis of initiation at this

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.

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

1	CGC	AUA	ACI	166	664	GCC	GCC	ecc	GCG	ALL	CGC	GC	CGC	CCC	CAG	CUI	CCG	cco	CCG	CAA	GATC	GG	cco	CTG	ccc	CAG	CCT	CCG	CGG	CAG	CCC	TGC	GIO	CACO	CAC
106	GGG	CCG	CGG	CCA	CCG	CCA	GCC	TGG	GGG	CCC.	ACC	TAC	ACT	ccc	CGC	AGT	GIG	CCC	CTG	CAC	CCCG	iса	TGT	AAC	CCG	GCC	AAC	ATC	CGG	CGA	GTG	TGC	CCT	CAG	ſAG
211	CII	CGG	CCC	CGG	GCT	GCG	ccc	ACC.	ACC	CAA	CAT	CAG	CTC	TCC.	AGCI	TCG	CAC	GTC	CGG	GAT	GGCA	GC	GGC	CAA	GGC	CGA	GAT	GCA	ATT	GAT	GTC	тсс	GCT	GCAG	àΑT
316	CTC	TGA	ccc	GIT	CGG	CTC	стт	TCC	TCA	стс	ACC	CAC	CAT	6GA	CAA	CTA	CCC	CAA	ACT	GGA	GGAG	AT	GAT	GCT	GCT	GAG	CAA	cGG	GGC	TCC	CCA	GTT	CCT	CGG	GC
1													н	D	N	Y	Ρ	к	1	F	F	н	м	1	1	s	N	6	A	Р	0	F	1	6	A
														-							-												-	-	
421	TCC	~~~					~ • ~																												
421	TGC .	caa	-	-	A0A	600	с нь		uge	LAA	TAA	LAG	LAG	LAG	LAG	LAG		LAG	JAG	LAG	LOGG	66	LGP	166	199	366	.66	LAG	LAA	CAG	CGG	LAG	LAG	LIGUI	
24	A	6	1	Ρ	E	G	S	6	G	N	Ν.	Ş	S	S	Ş	s	S	\$	s	s	G	G	G	G	G	G	G	s	N	s	G	S	5	Α	F
										•	• !		_	-												_	-								
526	CAA	TCC	TCA	AGG	6GA	GCC	GAG	CGA	ACA	ACC	CTA	CGA	GCA	CCT	GACO	CAC	AGA	GTC	CTT	TTC	TGAC	AT	CGC	TCT	GAA'	raa!	CGA	GAA	GGC	GCT	GGT	GGA	GAC	AAGI	TA
59	N	Ρ	0	G	E	Ρ	s	E	0	Р	Y	F	н	1	т	т	F	s	F	s	Ð	1	A	1	N	N	F	κ	Α		v	F	т	s	¥.
																								-						-			-		-
621	TCC	~ • ~	~~.																																
0.01	100	CAG		-	- NAC	u		ouu	-		- HCI		3	100	-			uun .	-	-	- GCA			CAU.	1001				010	-	1040	4004		-	.40
94	<u> </u>	2	Q		ب_	к	L	٢	P	1	1	11	υ	G	ĸ	ŀ	5	L	Ł	Р	A	P	N	S	G	N	1	ι	w	Ρ	Ł	P	Ľ.,	۲.	5
736	CCT	AGT	CAG	TGG	CCT	TGT	GAG	CAT	GAC	CAA	CCC.	ICC!	AC)	CTC	TTC/	ATC	CTC	AGC	SCC	TTC	TCCA	GC.	I GC	TTC.	ATCO	atc'	TTC	CTC	TGC	CTC	CCA	GAG	CCC.	ACCO	CT
129	L	۷	s	G	ι	۷	s	М	T	Ν	Ρ	Ρ	T	s	s	s	s	Α	Ρ	s	Ρ	А	A	s	s	s	s	s	А	s	Q	s	Ρ	Ρ	L
841	GAG	стб	TGC	CGT	228	GTC	саа	r G A	r a G	CAG	TCCI	ΔT	T AI	CTC.	4601	rac	٥٢٢	r a r i	٦T	TCC	таст	cci	- 0 01	rar'	IGAN	ΔΤ.	гтт	TCC	TGA	600	CCAL	4460	CA1	3600	TT
164		°.		u.	0	c.	ы.	0000	c	°.	000			°.			0	T	-	n	T	n .		- T	0	1		D	5	n .	0.00	¢	0		e .
104	3		^		,	3	•	0	3	3	٢		۰.	3	~	~	Ľ	<u> </u>	۴.	r	<u> </u>	٢	n	<u>`</u>	0	-	-	·	5	-r	4	3	v	~	۰.
946	TCC	TGG	CIC	TGC	AGG	CAC	AGC	CID	5CA	GTA	0000	PCC.	тсс	TGC	CTAC	ccc	TGC	CACO	CAA	666	IGGI	TTO	CCA	GGT	rcco	CATO	SAT	CCC	TGA	CTA	TCTO	STT:	TCC	ACAA	CA
199	Ρ	G	s	Α	G	T	A	L	Q	Y	Ρ	Ρ	Ρ	Α	Y	Ρ	Α	Т	κ	G	G	F	Q	۷	Ρ	м	I	Ρ	D	Y	L	F	Ρ	Q	Q
1051	ACA	666	AGA	сст	GAG	сст	GGG	CAC	ccc.	AGA	CCAG	SAAI	SCC	стт	CCAG	56G	TCT	GGA	SAA	CCG	тасс	CAG	5CAI	GCC.	TTC	SCTO	CAC	TCC	ACT	ATC	CACI	TAT	CAA	AGCO	TT
234	0	6	D	1	s	i.	G	т	Р	n	0	ĸ	Р	F	0	6	1	F	N	R	т	0	0	P	s	1	т	P	1	s	т	T	κ	A	F
											4								٠						°.					×.					
1166	000	c . c	TCA	ore								ICT.									CATC.									~~~~				-	
1130	· · ·	UNC.	0		000		CCA 0	004							UNCO		CCA		-UR				1001	CAG			200					- CON	acci	AGC	
269	A		Q	5	G	5	ų	U	L	ĸ	А	L		N		τ.	Q	5	Q	L.	1	ĸ	٢	2	к	м	к	ĸ	T	P	N	к	P	2	ĸ
													•																						
1261	GAC	ACC	ccc	CCA	TGA	ACG	ccc	GTA	TGC	TTG	CCC.	GT	(GA)	STC	CTGC	CGA	TCG	CCG	CT T	t t C	TCGC	TCO	6GA	TGA	SCT	(AC)	ACG	CCA	CAT	CCG	CATO	CCA	TAC	AGGC	CA.
304	T	ρ	Ρ	н	Εİ	R	Ρ	Y	Α	С	P	٧	E	S	C	D	R	R	F	s	R	s	D	ε	ι	T	R	н	I	R	1	н	T	G	0
						-								_			_		-	-															-
1366	GAA	000	стт	CCA	GIG	TCG	AAT	CTG	CAT	606	TAA	ITT	CAG	ICG	TAGI	16A	CCA	CCT	TAC	CAC	CAC	AT	206	CAC	CA	AC	166	CGA	GAA	000	TTT	IGCO	CTG	IGAC	ΔT
339	Tr	P	F	0	C	P	1	C	H	R	N	F	s	R	s	n	н		T	T	H	I	P	T	H	T	6	F	Fr	P	F	4	ř	n	T
	Ë	<u> </u>	÷	<u> </u>	<u> </u>		-	<u> </u>				·	÷			-	<u></u>		-		<u></u>	-	-	÷	<u> </u>	<u> </u>	-	-	<u> </u>	<u> </u>	÷	<u>.</u>	-	-	÷
	7.70	TCC				Tee											T												***						~
14/1	110	100	GH0			- int	CAG	040	TUA	194	ACG.		240	ac A	TACI		-	CUR		-010	10,40		-		300	•au	- 04		<i>м</i> чо	1011	Lun	300,0		-10	uuc.
374	c	G	R	<u></u>	+	A	к	s	0	<u>د</u>	R	ĸ	к	н	1	ĸ	1	н	L	R	9	ĸ	U	ĸ	ĸ	A	D	ĸ	s	۷	۷	A	S	S	A
1576	TGC	CTC	TTC	CCT	CIC	TTC	CTA	CCC.	ATC	CCC	AGTO	SGC (TACI	CTC	CTAC	CCC	ATC	ccco	CGC	CAC	CACC	TC/	ATT:	TCC	ATCO	:00/	\GT	GCC	CAC	CTC.	TTAC	CTCC	CTC	TCC6	GG
409	Α	s	s	L	s	s	Y	Ρ	s	Ρ	٧	Α,	Т	s	Y	Ρ	s	Ρ	А	т,	Т	s	F	Ρ	s	ρ	٧	Ρ,	Т	s	Y	s	s	P	G
									-					_									-						-						-
1691	CTO	стc	тас	ста	000	010	TCC	101	ana	CAG	IGGI	TT	222	ΔΤΟ	000	° TC	66 T	6601	-40	n a n	°TAT	601	101	Tan	202	scc.	rac	TTT		TGCI	CC AF	36 T/	° 461	- 400	тт
		e	т	v	0	e	P		ы	c	6			c.	D.	c	v		т	T	¥		c	v	0	0			D.		0	v	c	т	
***	<u>_</u>	2	-	-	-	-		^		-	<u> </u>		-		<u> </u>	5	-	- 1	÷		-	<u> </u>	-	-	-	<u></u>	^	·	'	^	Υ.	۰.	-	<u> </u>	<u>_</u>
1786	CCA	GTC	TGC	AGG	GGT	CAG	CAA	стс	CTT	CAG	CACO	CTC	140	GGG	TCTI	ттс	AGA	CATO	SAC.	AGC/	ACC	Ш	TICI	ICC.	rage	iac/	ΜĪ	IGA	AAT	LIEC	CTAP	AGG	3GA/	TGA	AA
479	Q	s	Α	G	۷,	s	N	s	F	s	τ	s	Т	G	ι	s	D	м	T	Α	τı	F	s	Ρ	R	T	I	ε	I	С	END)			
					~																														
1891	646	AGC	AAA	<u> </u>	AGG	66A	GCG	CGA	GAG.	ACA	A TAZ	AGO	AC.	466	AGGO	5AA	GAA	ATGO	SCC	CGC	AGA	660	GCI	TGCO	CTCI	TAG	GTO	CAG	ATG	GAAG	GATO	CTCA	AGAG	CCA	AG
	TCC	TTO		***	ст.										ю. т.т		COL			COT					rect		CA		- 46	CTO	CCTO				cc.
1330	TUU	IIC.	1.40	ICA			daci			acci	4007			110		I AG			GC			70										1000		1004	
2101	ICC	AAG	ιιc	Π¢	ALL	ICI.	AIU	. 44	166	aC I	GAI		a CA	166		166	AIA	AALL	.81	110	ROCH	10	uu.	icci	acco	(CAI	aci	-190	acci	2116	3010		inc,	IGC M	U.
2206	AGA	ACA	TCA	AGT	TGG	CTG	AAA	444	AA.	ATGO	GIC	TG	GCO	CTI	CAGA	AC	CCT	GCCC	; 1G	TAT	21110	GTA	ACA0	SCA	ICLE	GIGC	CA	IGG/	ATT	IIG		CCI	IIGe	6661	AI
2311	TCT	TGA	TGT	GAA	GAT.	AAT	TTG	CAT	AC TO	CTA.	TGI	ACI	TAT:	ITG	GAGT	TA	AA T	TCTO	CAC	ITT	566G	GAC	6660	GGA(SCA/	AGO	CA	AGC/	AAA	CA/	ATGO	STG/	ATCO	CICI	AT
2416	TIT	GTG	ATG	ATC	CTG	CTG	TGA	CAT	TAG	STT:	rga/	ACT	ITT:	TT	ITTI	ITT	TTG	AAGO	AG	CAG	ICCT.	AGO	STAT	T T AJ	ACTO	GAG	6CA	IGTO	GTC.	AGAG	GTGT	TGT	гтсо	GTT	AA
2521	TIT	TGT	6 6 A	тас	TGC	ICG	ACTO	STA4	сто	TCA	ICA1	GTO	AC.		TAC	266	TTT	STT	66	TIGO	GIT	TTI	ITGI	TTG	ITT	TGA	IAAJ	AAA /	AAA	ITTI	TTT	ITT	rgco	CGT	CC.
2626	CIT	166	177	CAA		117	CAC	STC	TGO	TGO	CTT	TGT	GT	AC.	CAC	CT.	TGC	CGAT	GGI	160	ACA	161	604	AATO	GTO	AGO	GG	ACA	GC	TCAC	сто	TAP	- 	TAA	GG
. 320	0001	TAC		TOP	TCT.		LCC-	20 4	601		0.0	0.01	GAT	0.00	CA +	C.F.		10	60	CA.		 16/	117	100	100	ATC	TA	164	101		ATTT		AC?		ΔT
:/31	aaa	1.460	2410	AUR	101		100	3040	au		ant		-04	uni(NAMER	NON		- 104	100	ani(1011	100	0.11		-404	007	100	-uru	- 10/0		111	407			Te
:836	CTG/	AC]	1010	UAA.	AAG'	CD	00	um	IN	AC TO	AAA	A10	IAC	ATT	TAT	¢¢.	A 1 G	100	itili)	1G11	GGA	A16	IC 16	aCG(114		AC	I GAL	1 AI	3000	3016	AU		01A	16
2941	CTAT	1GAJ	ACA.	TGA.	AGT	ICA.	ITA.	IIII	GTO	GTI	TTA	TTI	TAC	TT	GTA	CT	TGT	STTI	GC.	ITA	IACA.	AAG	iTG/	ACTI	GTI	fGG	ICT.	IATA	444	:AC/	ALL6	iAA1	GCG	CIT	ľA
8046	CTGO	:00/	A TGO	66A	TATO	GTG	STGI	IGT/	TCC	TTC	AGA	AAA	ATI	AA	IAGG	iAA.	AA T <i>i</i>	AAG	iAAJ	ACT/	ACTI	GG1	AA	ww	١										

Α		в	Residues	Sequence
TFIIIA Consens	us KPYVCXXDGCDKRFTKKXXLKRHXXXHTGE	Rl	91-98	ΤSΥΡЅQΤΤ
NGFI-A Repeat	R P Y A C P V E S C D R R F S R S D E L T R H I R I H T G Q	R2 R3	180-187 188-195	DIFPEPQS
NGFI-A Repeat	K P F Q C R I C M R N F S R S D H L T T H I R T H T G E	R4 R5	413-420 421-428	S S Y P S P V A T S Y P S P A T
Kruppel	K P F D C K I C S R S F G Y H V - L Q N H E R T H T G E	R6	429-436	TSFPSPVP
ADRI	RSFVCEV CTRAFARQEHLKRHYRSHINE	R8	445-452	STYPSPAH
		R9 R10	453-461 462-468	S G F P S P S V T T Y A S V P P
		R11	476-483	STFQSAGV
		•		[т][т][ғ] рсруу

Consensus sequence $\begin{bmatrix} T & T & F \\ S & S & Y \end{bmatrix}$ P S P X X

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.

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 NH2terminal region of the molecule (Fig. 2). These repeats are aligned with each other, along with a consensus sequence, in Fig. 3B.

The NH₂-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 Olinked, 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. 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)
- 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).
- 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 S. Whittemore et al., Proc. Natl. Acad. Sci. U.S.A. 83, 817 (1986); M. Taniuchi, J. B. Schweit-zer, E. M. Johnson, Jr., *ibid.*, p. 1950.
 H. Thoenen, Y.-A. Barde, D. Edgar, in Neuroscere-tion and Brain Peptides, vol. 28 of Advances in Divident Performance in Neuroscenes in Neur
- Biochemical Psychopharmacology, J. B. Martin, S. Reichlin, K. L. Bick, Eds. (Raven, New York, 1981), pp. 263–272.
- 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)
- U. B. Rosenberg et al., *ibid.* 319, 336 (1986).
 U. B. Rosenberg et al., *ibid.* 319, 336 (1986).
 J. Greene and A. Tischler, in Advances in Cellular Neurobiology, S. Federoff and L. Hertz, Eds. (Aca-demic Press, New York, 1982), vol. 3, pp. 373– U. M. Statistical Science Scienc 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).
 L. R. Karns, S.-C. Ng, J. A. Freeman, M. C. Fishman, *Science* 236, 597 (1987). 12.

- 13. T. Curran and J. I. Morgan, ibid. 229, 1265 (1985); M. Greenberg, L. Greene, E. Ziff, J. Biol. Chem.
 Z60, 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).
 K. Kelly, B. Cochran, C. Stiles, P. Leder, Cell 35, Construction.
- 14.
- 603 (1983).
- 5. 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 following apple/dx1, DNA use collected by two press. 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 mM Tris, pH 8.3 at 25°C 140 mM potassium acetate, 10 mM magnesium acetate, oligo(dT) (80 μ g/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 Agt10 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).
- C. Setoyama, R. Frunzio, G. Liau, M. Mudryj, B. de 17 Crombrugghe, Proc. Natl. Acad. Sci. U.S.A. 83, 3213 (1986).
- J. Milbrandt, unpublished data 18.
- M. Kozak, Cell 44, 283 (1986).
 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). S. R. Haynes et al., Proc. Natl. Acad. Sci. U.S.A. 84,
- 23. 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 mm²), and four replica filters were prepared from each plate. Probes were prepared by synthesizing singlestranded cDNA from 2 μ g 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 μ Ci of ³²P-labeled deoxycytidine 5'-triphosphate were included. The RNA template was removed by incubation at 68°C for 1 hour with 50 mM 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 $0.2 \times$ standard saline citrate and 0.1% SDS at 65°C. Plaques that produced a differential signal were identified by inspection and picked for condary screening.
- G. Helms, M. Graham, J. Dutchik, M. Olson, DNA 26. (New York) 4, 39 (1985)
- A. Feinberg and B. Vogelstein, Anal. Biochem. 67, 27. 15 (1984).
- 28. F. Sanger, S. Nicklen, A. R. Coulson, Proc. Natl. Acad. Sci. U.S.A. 74, 5463 (1977).
- S. Henikoff, Gene (Amsterdam) 28, 351 (1984) 29.
- 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 Neurologi cal and Communicative Diseases and Stroke.

4 June 1987; accepted 28 August 1987