Purification, Characterization, and Partial Amino Acid Sequence of Nerve Growth Factor from Cobra Venom†

Ruth A. Hogue-Angeletti,¶ William A. Frazier,‡ John W. Jacobs, Hugh D. Niall,§ and Ralph A. Bradshaw*,#

ABSTRACT: The nerve growth factor (NGF) from *Naja naja* (cobra) venom has been purified and its structure compared to the NGF from mouse submaxillary gland. A two-step purification procedure has been devised, consisting of a gel filtration step in 1 *M* acetic acid followed by chromatography of the active pool on carboxymethylcellulose at pH 5. The molecular weight of the native protein was found to be 28000, and this value was reduced by approximately one-half under denaturing conditions. These values are comparable to those obtained for mouse 2.5S or βNGF. Tryptic peptide maps of *S*-[^{14}C]carboxymethyl NGF gave the number of labeled peptides expected for a structure composed of two identical or very similar subunits. Thus, the quaternary structures of mouse and cobra NGF are the same. Cyanogen bromide (CNBr) treatment of *Naja naja* NGF produced three fragments, of which two were purified to homogeneity. These fragments and the whole protein were analyzed in the automated protein Sequencer. The amino-terminal CNBr fragment of the protein was also subjected to digestion by thermolysin and the resultant peptides were purified and characterized. These data plus those from the characterization of the tryptic peptides provided the basis of the construction of a tentative primary structure of *Naja naja* NGF which is approximately 60% identical with mouse NGF.

Nerve growth factor (NGF)[1] is a pleiotypic effector of the differentiation and maintenance of the sympathetic and sensory neurons of the peripheral nervous system of vertebrates (Frazier et al., 1972; Boyd et al., 1974). Extensive structural studies on NGF purified from male mouse submaxillary glands indicated it to be distantly related in amino acid sequence and in certain features of three-dimensional structure to insulin (proinsulin) (Frazier et al., 1972, 1973). These structural comparisons are further borne out by the striking similarities in their biological effects and receptor binding properties (Frazier et al., 1974a,b). In order to extend the basis for evolutionary comparisons, it was of interest to compare an NGF protein from the more primitive class Reptilia (Levi-Montalcini and Cohen, 1956; Angeletti, 1970, 1971) with that of both mouse NGF and the insulins (proinsulins).

The nerve growth promoting activity of the venom from the cobra, *Naja naja,* has been isolated as a homogeneous protein and the subunit structure and partial amino acid sequence determined. From the data presented in this article, it is clear that the structure of cobra venom NGF is closely related to that of the mouse protein in the same manner established for other protein families whose members are evolutionarily related. It is, however, not more closely related to the insulin family than the mouse protein.

Experimental Procedure

Materials. Lyophilized venom from the Indian cobra, *Naja naja,* was obtained from the Miami Serpentarium. Materials for chromatography and gel filtration were obtained from Sigma Chemical Co. (St. Louis, Mo.). Trypsin, treated with (1-tosylamido-2-phenyl)ethyl chloromethyl ketone (Tos-PheCH$_2$Cl)[1] was obtained from Worthington Biochemical Corp. Thermolysin was obtained from Daiwa Kasei K.K., Osaka, Japan. Guanidine-HCl was the Ultra-Pure grade of Schwarz/Mann Biochemicals. Pyridine was redistilled after the addition of solid ninhydrin (1 g/l.) to the distillation flask. Crystalline iodo[1-^{14}C]acetic acid with a specific activity of 15 mCi/mmol was obtained from New England Nuclear Corp. Unlabeled iodoacetic acid was recrystallized from cold petroleum ether (bp 30-60°) before use.

Methods. For the first step in purification of *Naja naja* NGF, 3 g of lyophilized venom (a total of 40 g of venom was processed) was dissolved in 15 ml of 1 *M* acetic acid and applied to a column of Sephadex G-50 fine, 5 × 250 cm, at 4°. The column was eluted at 30 ml/hr with 1 *M* acetic acid and the fractions assayed for absorbance at 280 nm. After pooling, each fraction was dialyzed against three changes of 100 vol each of 50 m*M* sodium acetate (pH 5.0), 4°. Any precipitate formed during dialysis was discarded after centrifugation at 5000*g*. The soluble fraction was applied to 1.5 × 25 cm columns of carboxymethylcellulose

† From the Institute for Cancer Research, Philadelphia, Pennsylvania 19111, the Department of Biological Chemistry, Division of Biology and Biomedical Sciences, Washington University School of Medicine, St. Louis, Missouri 63110, and the Endocrine Unit, Massachusetts General Hospital and the Harvard University School of Medicine, Boston, Massachusetts 02114. *Received June 6, 1975.* Supported by research grants from the U.S. Public Health Service, NS-10229, CA-06927, and RR-05539 from the National Institutes of Health, and by an appropriation from the Commonwealth of Pennsylvania.

¶ Present address: Division of Neuropathology, University of Pennsylvania, Philadelphia, Pa. 19174.

‡ Present address: Department of Psychiatry, University of California at San Diego, La Jolla, Calif. 92037.

§ Present address: The Howard Florey Institute of Experimental Physiology and Medicine, University of Melbourne, Parkville, Victoria, 3052 Australia.

U.S. Public Health Service Research Career Development Awardee AM-23968. To whom correspondence should be addressed at Washington University, St. Louis, Mo.

[1] Abbreviations used are: NGF, nerve growth factor; Tos-PheCH$_2$Cl, (1-tosylamido-2-phenyl)ethyl chloromethyl ketone; Tp, tryptic; Th, thermolytic; CNBr, cyanogen bromide; S-CM, *S*-carboxymethyl; CM, carboxymethyl; Pth, phenylthiohydantoin.

(CM-[1]) equilibrated with the dialysis buffer. After adequate washing with the starting buffer, a three-chambered gradient from 0 to 0.7 *N* NaCl in starting buffer was used to elute the protein at 40 ml/hr. Analytical and preparative polyacrylamide gel electrophoresis were performed on 6 cm, 7.5% gels at pH 9.5 and 18° using the system of Davies (1964) and Ornstein (1964).

The bioassay for NGF was performed using a modification of the method described by Levi-Montalcini and Angeletti (1968).

Molecular weight of the native NGF samples was determined using a Spinco Model E ultracentrifuge, following the high-speed technique of Yphantis (1964). Samples were dissolved in 0.1 *M* sodium acetate (pH 5.0) and dialyzed overnight before each analysis. Analyses were performed at 20° and 45000 rpm, and monitored using an ultraviolet scanner. The molecular weight of the reduced and alkylated NGF was calculated from sodium dodecyl sulfate–polyacrylamide gel electrophoresis (Weber and Osborn, 1969).

Radioactive and nonradioactive *S*-carboxymethyl *Naja naja* NGF were prepared as previously described for mouse NGF (Angeletti et al., 1971). Radioactivity was measured in a Packard Model 3370 liquid scintillation spectrometer with a counting efficiency for ^{14}C of 84%. Aliquots for analysis were dissolved in 10 ml of Bray's water miscible scintillation liquid. Radioactivity in column chromatography experiments was measured with 50-μl aliquots withdrawn from alternate fractions.

Tryptic digestion was carried out using Tos-PheCH_2Cl-treated trypsin added in a ratio of 1:100 to the NGF on a weight basis according to Bradshaw et al. (1969). Two micromoles of *S*-[^{14}C]carboxymethyl *Naja naja* NGF was used in the preparation of the tryptic fingerprint. The reaction mixture was maintained at pH 8.8, 37° by means of a Radiometer pH stat, Model T-11. The insoluble material remaining after acidification to pH 2 with *N* HCl was removed by centrifugation.

The soluble tryptic (Tp) peptides were fractionated on a 0.9 × 20 cm column of Dowex 50-X8 equilibrated in 0.05 *N* pyridine acetate (pH 2.5) and eluted with a double linear gradient of pyridine acetate buffers as described by Bradshaw et al. (1969). The separation of peptides was monitored with a Technicon Autoanalyzer equipped for automatic alkaline hydrolysis and ninhydrin analysis (Hill and Delaney, 1967). The isolated peptides were recovered after removal of the solvent by rotary evaporation.

Fragmentation of *S*-carboxymethyl *Naja naja* NGF (1 μmol) with cyanogen bromide was performed at room temperature in 70% formic acid with a 20-fold molar excess of cyanogen bromide for 24 hr in a nitrogen atmosphere (Gross and Witkop, 1961). The cyanogen bromide was blown off with a nitrogen stream, and the sample was brought to near dryness by rotary evaporation. It was dissolved in 4 ml of 0.05 *N* pyridine acetate (pH 2.5) and applied to a 2 × 125 cm column of Sephadex G-75 equilibrated in the same buffer at 10 ml/hr at 25°. Subsequent fractionation was performed on a column of Sephadex G-50 fine, 2 × 125 cm, equilibrated and eluted under the same conditions. The profile was detected by following the absorbance at 280 nm.

Thermolytic digestion of fragment CB II-3 was performed at 37° as described previously (Angeletti et al., 1973a,b) with a final concentration of enzyme to substrate of 0.5%. The reaction was monitored in the pH stat at pH 8.0 for 20 hr and terminated by acidification to pH 2 with 6 *N* HCl. The soluble peptides were fractionated on Dowex 50-X8 according to the procedure described for the tryptic peptides and monitored automatically with ninhydrin after alkaline hydrolysis as described above.

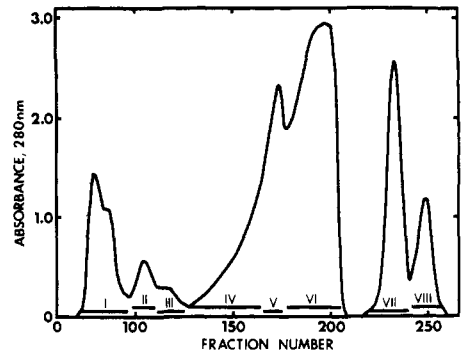

FIGURE 1: Gel filtration of crude *Naja naja* venom on Sephadex G-50 at 4°. Bars and Roman numerals indicate the pools made from the 30-ml fractions.

The procedures for isolating and characterizing both tryptic and thermolytic peptides on Dowex 1-X2 and Sephadex G-25 were essentially those described previously (Angeletti et al., 1973a,b). A few thermolytic peptides were purified using preparative thin-layer electrophoresis at pH 4.7.

Amino acid analyses were performed on a Beckman 120C or a Durrum D-500 automatic amino acid analyzer. The tryptophan content of the native protein was determined spectrophotometrically by the method of Edelhoch (1967). Half-cystine was determined as cysteic acid (Moore, 1963) and as *S*-carboxymethylcysteine on the amino acid analyzer.

Analyses of amino-terminal sequences of whole *S*-carboxymethyl NGF and the cyanogen bromide fragments thereof were performed on the Beckman Model 890C Automatic Sequencer according to Jacobs and Niall (1975).

Sequential degradation of purified peptides was carried out according to Edman (1956). The modifications of Konigsberg and Hill (1962) were adopted to facilitate the use of subtractive sequence analysis on the amino acid analyzer. In some cases the sequence was determined using the dansylation technique of Hartley (1970) and Gray and Smith (1970). Dansylated amino acid derivatives were identified by chromatography on polyamide sheets by a modification of the procedure described by Woods and Wang (1967) (Angeletti et al., 1973b). Carboxyl-terminal sequences of several peptides were determined using carboxypeptidases A and B.

Results

Purification of NGF from Naja naja Venom. Due to the known stability of the biological activity of mouse NGF at acidic pH, 1 *M* acetic acid was chosen as the solvent for the initial fractionation by gel filtration of crude *Naja naja* venom. Fractions from this separation, shown in the elution profile in Figure 1, were assayed for NGF activity using the in vitro bioassay. However, interference in the bioassay system caused by toxins and other substances present in the fractions was so great as to preclude unambiguous identification of NGF activity in any area of the profile. Therefore, the eight pools of the gel filtration column were dialyzed at 4° overnight against 50 m*M* sodium acetate buffer (pH 5.0) and fractionated on a column of CM-cellulose. Bioassays performed on individual fractions of each column re-

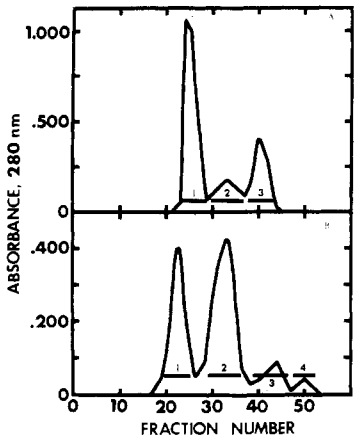

FIGURE 2. Chromatography of Sephadex G-50 pools on CM-cellulose at 4°. Bars and Arabic numerals show the pool made of the 10-ml fractions: (A) subfractionation of G-50 pool II; (B) subfractionation of G-50 pool III.

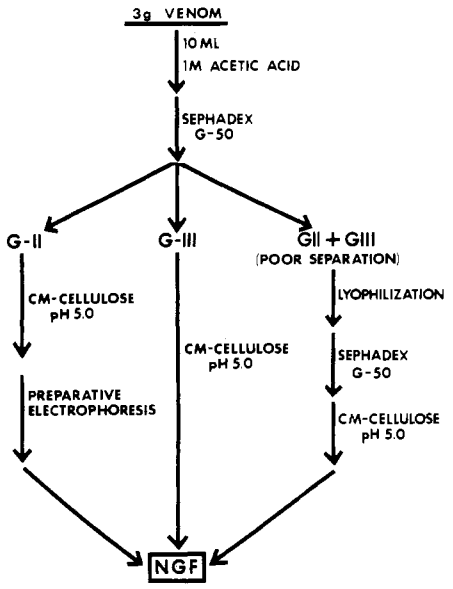

FIGURE 3: Flow diagram of alternative routes of purification of *Naja naja* NGF.

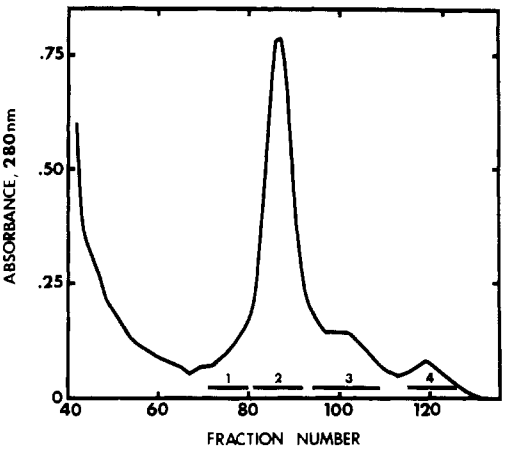

FIGURE 4: Preparative polyacrylamide gel electrophoresis of pool II-2 of *Naja naja* NGF. The gel was a 6 cm, 7.5% acrylamide gel at pH 9.5, 18°. The current was slowly increased to 15 mA after application of the sample. The flow rate was 12 ml/hr. One-milliliter fractions were collected. The first unmarked peak is the position of elution of the tracking dye. Bars and Arabic numerals indicate the pools made.

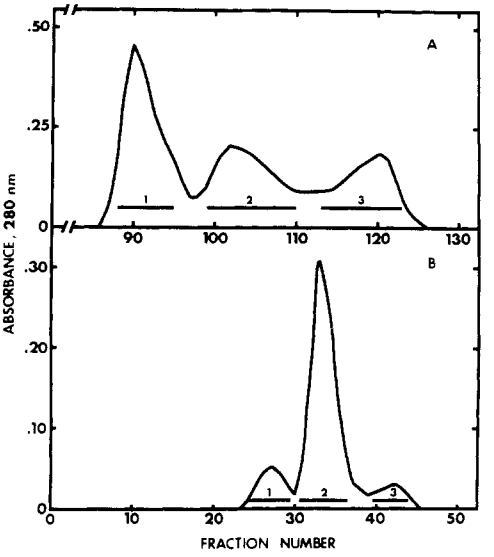

FIGURE 5: Alternative method of fractionation when pool G-II and G-III were not well separated: (A) refractionation of the combined pools on G-50 fine in 1 *M* acetic acid under the same conditions as in Figure 1; (B) CM-cellulose chromatography of pool G-II/III-3 using the same conditions as described in Figure 2.

vealed high levels of NGF activity in the CM-cellulose fractionations of only pools G-II and G-III, whose elution profiles are shown in Figure 2. In each separation, pool 2 indicated in the figure contained the NGF activity.

Analytical polyacrylamide disc gel electrophoresis at pH 9.5 revealed a single band for pool 2 of the CM-cellulose column of G-III. Contiguous sections of an unstained gel were homogenized in physiological saline and centrifuged at 1000 rpm in a Sorvall GLC-2 centrifuge and the supernatants assayed for NGF activity. Biological activity was localized at the same position as the protein band. However, pool 2 from G-II also possessed three additional minor bands as judged from the protein stain.

Figure 3 is a flow diagram of the purification procedure used. When pooled as indicated in Figure 2B, only two purification steps were necessary to obtain the majority of NGF in the venom in pure form. Since the amount of NGF present in pool 2 of G-II was not insignificant, attempts were made to further purify this material. The only successful procedure was preparative polyacrylamide gel electrophoresis, using the same conditions as those applied for analytical procedures. Figure 4 shows the elution profile obtained from the gel. The major peak (2) possessed NGF activity and was homogeneous as judged by analytical electrophoretic procedures. The other three peaks (1, 3, and 4) corresponded to the three minor electrophoretic bands present in the fraction applied. When aberrations in the gel filtration column caused the G-II and G-III peak and shoulder to be poorly resolved, CM-cellulose chromatography did not yield homogeneous material. In this case, the active CM-cellulose pools could be purified to homogeneity using the preparative electrophoretic technique or, alternatively, as shown in Figure 3, the original G-II and G-III pools could be combined, directly lyophilized, and re-applied to the same Sephadex G-50 column. The superior resolution obtained in this case (Figure 5) permitted the ready purification of NGF on the CM-cellulose column.

A total of 40 g of *Naja naja* venom was fractionated according to the above procedures. An average yield of 1.5 mg of purified NGF was obtained per g of venom.

Table I: Amino Acid Composition of *Naja naja* NGF.

Amino Acid	Residues/Subunit (Mol Wt 13000) 24 hr	48 hr	72 hr	Av	Integral
Aspartic acid	16.02	16.01	16.03	16.02	16
Threonine	12.35	11.97	11.96	13.00[a]	13
Serine	7.68	7.02	6.87	8.00[a]	8
Glutamic acid	10.38	9.92	10.02	10.10	10
Proline	5.01	5.60	5.23	5.28	5
Glycine	5.91	5.74	5.63	5.76	6
Alanine	4.88	4.68	4.65	4.74	5
Half-cystine				6.00[b]	6
Valine	8.94	10.51	9.72	9.72[c]	10
Methionine	2.03	1.57	1.71	1.77	2
Isoleucine	4.58	5.68	5.66	5.66[c]	6
Leucine	2.99	3.33	2.81	3.04	3
Tyrosine	2.73	2.82	2.76	2.77	3
Phenylalanine	3.81	3.68	3.85	3.78	4
Lysine	9.60	9.50	9.50	9.53	10
Histidine	3.63	3.52	3.58	3.58	4
Arginine	2.85	2.64	2.85	2.78	3
Tryptophan				3.00[d]	3
				Total	117

[a] Value extrapolated to zero time of hydrolysis. [b] Determined as cysteic acid (Moore, 1963) and as *S*-carboxymethylcysteine. [c] Value at the longest hydrolysis time was used. [d] Estimated by the spectrophotometric method of Edelhoch (1967).

Molecular Weight of Naja naja NGF. The molecular weight of *Naja naja* NGF was measured using the sedimentation equilibrium method. At pH 5.0 in 0.1 *M* sodium acetate buffer, the molecular weight was calculated to be 28000 ± 1000. By the same method and using the same conditions, mouse NGF had been found to have an apparent molecular weight of 29000 ± 1000. Analysis of the molecular weight of reduced, carboxymethylated *Naja naja* NGF on 10% sodium dodecyl sulfate–polyacrylamide gels gave a molecular weight of 13000. Estimates of the subunit molecular weight of mouse NGF by this method range from 12500 to 14000.

Amino Acid Composition of Naja naja NGF. The amino acid composition of cobra NGF (Table I) was calculated from acid hydrolysates incubated for 24, 48, and 72 hr at 110°. A minimum of two analyses was calculated to give the final values for each time point. Previous work (Angeletti, 1970) suggests that the six half-cystinyl residues per subunit are all present in disulfide linkage.

Tryptic Digest of S-[^{14}C]Carboxymethyl NGF. As a first step in comparing the primary structures of mouse and cobra venom NGF's, a tryptic hydrolysis of *S*-[^{14}C]carboxymethyl *Naja naja* NGF was performed. On the basis of the amino acid composition of the cobra venom NGF, approximately 14 tryptic peptides are to be expected if the subunits are indeed identical.

The elution profile for fractionation of the soluble tryptic peptides on the ion-exchange column is shown in Figure 6. Amino acid compositions of corresponding pools, after the indicated subsequent purification, are shown in Table II. There are four major radioactive pools. Pool I contained no peptidic material, as judged by amino acid analysis, and was presumed to be *S*-[^{14}C]carboxymethyl-2-mercaptoethanol incompletely removed by dialysis from the alkylation reaction. Pool Tp V,[1] however, did contain a peptide apparently possessing two residues of CM-cysteine, whereas pool Tp X, upon subsequent purification, yielded three peptides containing CM-cysteine (Tp X-1, Tp X-2, and Tp X-3). However, Tp X-1 and Tp X-2 are apparently derived from an incomplete tryptic cleavage. Pool Tp XII, the last remaining radioactive pool, contained a radioactive peptide in approximately 20–25 yield. Further attempts at purification of this pool still resulted in impure material in insufficient amounts for sequence analysis.

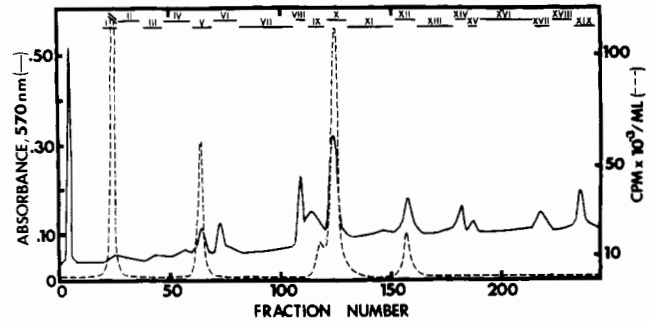

FIGURE 6: Elution profile of soluble tryptic peptides of *S*-[^{14}C]carboxymethyl *Naja naja* NGF on a 0.9 × 20 cm column of Dowex 50-X8 at 55°. The column was developed at 30 ml/hr with a double linear gradient of pyridine acetate, as described in the text, and monitored automatically by ninhydrin analysis (—) after alkaline hydrolysis. Fractions of 3.0 ml were collected. Radioactivity (- - -) was counted in 50-μl aliquots taken from alternate fractions. Fractions were pooled as indicated by the bars and corresponding Roman numerals.

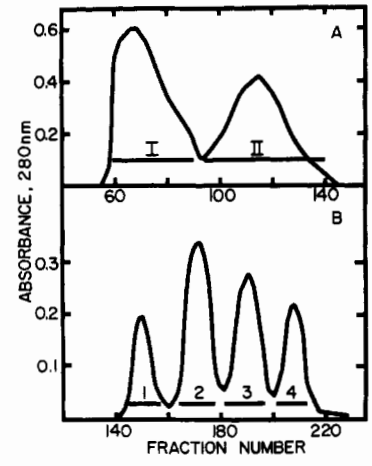

FIGURE 7: Separation of the cyanogen bromide fragments of *S*-carboxymethyl *Naja naja* NGF by gel filtration: (A) elution profile of complete reaction mixture on Sephadex G-75; bars and Roman numerals show the pools made of the 1-ml fractions; (B) elution profile of pool II (from A) on Sephadex G-50.

Pools Tp III and Tp IV contained a peptide of the same composition, probably due to deamidation of Tp IV. The yield of Tp III shown is the combined yield of Tp III plus Tp IV. Pools Tp VI and Tp VIII contained no significant peptidic material. All of the ninhydrin-positive pools containing peptidic material were analyzed by Edman degradation and carboxypeptidase digestion, as indicated later in Figure 8 (vide infra).

CNBr Fragmentation of Najanaja NGF. Cobra venom NGF has one additional residue of methionine as compared to mouse NGF, yielding three fragments after CNBr cleavage rather than the two obtained from the mouse protein (Angeletti et al., 1973a). The elution pattern of the fractionation of the CNBr fragments of *S*-carboxymethyl *Naja naja* NGF is shown in Figure 7A. The first peak appeared to be high molecular weight material, and was allowed to react a second time with CNBr before further subfractiona-

Table II: Amino Acid Composition of Tryptic Peptides Isolated from *S*-[^{14}C]Carboxymethyl *Naja naja* NGF.

Amino Acid	III	V	VIII	IX	X-1	X-2	X-3	XIV	XV	XVII	XIX
Aspartic acid			1.2 (1)	1.00 (1)	1.77 (2)	1.75 (2)	3.00 (3)		0.95 (1)		
Threonine	1.00 (1)	2.15 (2)	2.9 (3)	1.07 (1)			2.82 (3)				
Serine				0.95 (1)	0.93 (1)	0.78 (1)	2.82 (3)				
Glutamic acid	2.35 (2)	0.93 (1)		2.24 (2)	0.93 (1)	0.74 (1)	1.14 (1)				
Proline					2.70 (3)	3.21 (3)					
Glycine				1.10 (1)	1.19 (1)	1.33 (1)	1.17 (1)		1.01 (1)		
Alanine		1.06 (1)	1.2 (1)	2.13 (2)							
CM-cysteine		1.56 (2)			1.37 (2)	0.65 (1)	0.56 (1)				
Valine		1.76 (2)[a]								0.96 (1)	
Methionine				0.88 (1)							
Isoleucine		1.75 (2)	0.78 (1)				1.85 (2)				0.82 (1)
Leucine				0.99 (1)							
Tyrosine	0.74 (1)						0.78 (1)			0.82 (1)	
Phenylalanine	1.98 (2)						0.81 (1)				0.97 (1)
Lysine	1.00 (1)	1.00 (1)	0.83 (1)		1.00 (1)	1.00 (1)	0.67 (1)	(1)	0.97 (1)	1.15 (1)	
Histidine							0.80 (1)				
Arginine				0.87 (1)	0.72 (1)						1.03 (1)
Tryptophan[c]				(1)			(1)				
Total	7	11	7	12	12	10	19	1	3	3	3
Yield	70%	58%	74%	45%	33%	25%	64%	36%	49%	67%	52%
Purification procedure[b]	G-25	DXI	G-25	DXI	DXI	DXI	DXI				DXI

[a] 72-hr hydrolysis. [b] G-25, Sephadex G-25; DXI, Dowex 1-X2. [c] Determined qualitatively by spot test with *p*-dimethylaminobenzaldehyde.

Table III: Amino Acid Composition of Cyanogen Bromide Fragments of *S*-[^{14}C]Carboxymethyl *Naja naja* NGF.

	CB II-2		CB II-3	
Amino Acid	Residues/mol	Integral	Residues/mol	Integral
Aspartic acid	8.8	9	5.3	5
Threonine	4.6	5	5.0	5
Serine	4.0	4	3.3	3
Glutamic acid	5.9	6	2.2	2
Proline	3.3	3	2.3	2
Glycine	2.2	2	2.1	2
Alanine	1.1	1	1.8	2
CM-cysteine	2.6	3	1.2	1
Valine	1.9	2	4.6	5
(Methionine)[a]	(1)	(1)	(1)	(1)
Isoleucine	1.2	1	0.6	1
Leucine	2.0	2	2.1	2
Tyrosine	2.6	3		
Phenylalanine	2.7	3		
Lysine	5.7	6	2.2	2
Histidine	1.0	1	2.6	3
Arginine	1.0	1		
Tryptophan[b]	(1)	(1)	(1)	(1)
Total		54		37

[a] As homoserine. [b] Estimated by the spectrophotometric method of Edelhoch (1967).

tion. However, no smaller fragments were generated by this second treatment with CNBr. Pool II, after concentration by rotary evaporation, was applied to a column (2 × 125 cm) of Sephadex G-50, and the profile shown in Figure 7B obtained. Pool II-1 still contained unfragmented or aggregated material, as judged by amino acid analysis. Pools II-2 and II-3 appeared to be pure, and their amino acid compositions are shown in Table III. Further attempts to purify pool II-4 by a variety of methods did not yield a suitable product and no further structural analyses were attempted.

Automatic Sequencer Analysis of Naja naja NGF and

Table IV: Sequencer Analysis of Fragment CB II-2 of *Naja naja* NGF.

Cycle No.	Residue Identified	Yield (nmol)[a]
1	Glu	44.0
2	Asn	29.1
3	Val	40.0[b]
4	Asn	26.3
5	Leu	23.6[b]
6	Asp	18.4
7	Asn	10.3
8	Lys	10.3
9	Val	4.0[b]
10	Tyr	3.1
11	Lys	3.1
12	Gln	1.8
13	Tyr	3.2

[a] Calculated from phenyl[^{35}S]thiohydantoin recovery. [b] Calculated from gas–liquid chromatography.

Table V: Sequencer Analysis of Fragment CB II-3 of *Naja naja* NGF.

Cycle No.	Residue Identified	Yield (nmol)[a]
1	Glu	51.5
2	Asp	46.5
3	His	23.7
4	Pro	32.1
5	Val	35.6
6	His	18.6
7	Asn	21.4
8	Leu	23.2
9	Gly	15.4
10	Glu	12.4
11	His	4.1
12	Pro	10.2
13	Val	6.2

[a] Calculated from phenyl[^{35}S]thiohydantoin recovery.

Table VI: Amino Acid Composition of Thermolytic Peptides of CB II-3 of *S*-[^{14}C] Carboxymethyl *Naja naja* NGF.

Amino Acid	I	II-1	II-2	III	IV	V	VI	VII	VIII	IX
Aspartic acid	1.98 (2)	0.72 (1)			0.93 (1)			1.22 (1)	1.28 (1)	
Threonine	2.17 (2)						2.24 (2)	1.07 (1)		
Serine	0.68 (1)	0.75 (1)	0.68 (1)	0.89 (1)						
Glutamic acid					0.78 (1)	1.13 (1)				
Proline					1.16 (1)	0.83 (1)				
Glycine						1.16 (1)		1.10 (1)		1.00 (1)
Alanine			0.91 (1)							
CM-cysteine	0.44 (1)	0.48 (1)								
Valine	0.74 (1)	1.08 (1)		1.00 (1)			1.01 (1)		1.10 (1)	0.83 (1)
Isoleucine								1.03 (1)		
Leucine						0.98 (1)				
Histidine					0.94 (1)	1.31 (1)			0.83 (1)	
Lysine							1.15 (1)	1.04 (1)		1.29 (1)
Tryptophan										(1)
Total	7	4	2	2	4	5	4	5	3	4
Amino terminus	Val	Val	Ala	Val	Glu	Leu	Val	Thr	Val	Trp
Yield	25%	31%	35%	27%	46%	42%	22%	27%	23%	44%
Purification[a]	TLE	TLE	TLE			TLE				TLE

[a] TLE, thin-layer electrophoresis, pH 4.7.

CNBr Fragments, CB II-2 and CB II-3. Analyses of the N-terminal sequences of both *S*-carboxymethyl *Naja naja* NGF and the two pure CNBr fragments, CB II-2 and CB II-3, were performed with the automatic Sequencer. The degradation of the whole protein stopped after identification of the first two residues, Glu-Asp. The results for CB II-2 and CB II-3 are shown in Tables IV and V, respectively.

Thermolytic Cleavage of CB II-3. Fragment CB II-3 was digested with thermolysin and chromatographed on a column of Dowex 50-X8. The amino acid compositions of the isolated peptides are shown in Table VI. The amino-terminal residue of each peptide was determined using the dansyl technique, and is listed in the table.

Discussion

The goal of this study was to determine the nature of the molecular entity responsible for the nerve growth promoting activity of *Naja naja* venom and assess the extent of its relationship to mouse NGF and the class of related proteins, the proinsulins. In view of the expensive nature of the starting material, it was necessary to devise an efficient preparative procedure. The purification of NGF from the venom was hindered by the interference of toxic components with the bioassay, which necessitated the subfractionation of each G-50 pool on CM-cellulose columns before reliable bioassay results could be obtained. For this reason, it was not possible to calculate the absolute yields of NGF from the preparation.

The molecular weight of purified *Naja naja* NGF was determined by hydrodynamic measurements similar to those used for mouse NGF (Angeletti et al., 1971). The native molecular weights of both proteins were about 28000. In each case this value was reduced to half under denaturing conditions, suggesting the same subunit structure.

Comparison of the amino acid compositions of mouse NGF (Angeletti et al., 1971) and *Naja naja* NGF from these experiments and previous work (Angeletti, 1970) suggests structural similarity of the polypeptide chains. In Table VII are shown the integral values for the amino acids obtained in this study (column one), compared to the previously published values (column two), which were based on the native molecular weight. When these latter values are divided by two (column three), the differences are seen to be minimal. Furthermore, comparison to the amino acid composition based on the tentative primary structure (vide infra) shows the differences between the expected and derived compositions to be within experimental error.

In the last column of Table VII are listed the integral values for mouse NGF. As can be seen, the number of most residues are similar. The most notable differences occur in proline and the charged residues, the *Naja naja* protein possessing seven more acidic residues and three fewer basic residues than mouse NGF. This corresponds to the measured differences in isoelectric points of the mouse (9.4) and

Table VII: Comparison of the Amino Acid Compositions of *Naja naja* NGF and Mouse NGF.

	Residues/mol				
Amino Acid	*Naja naja* NGF (Present)	*Naja naja* NGF (Published)[a] Dimer	Monomer	*Naja naja* NGF (Sequence)	Mouse NGF[b]
Aspartic acid	16	34	17	16	11
Threonine	13	26	13	12	14
Serine	8	16	8	8	11
Glutamic acid	10	20	10	10	8
Proline	5	10	5	5	2
Glycine	6	16	8	7	5
Alanine	5	12	6	5	8
Half-cystine	6	8	4	6	6
Valine	10	18	9	10	13
Methionine	2	2	1	2	1
Isoleucine	6	12	6	6	5
Leucine	3	8	4	3	3
Tyrosine	3	4	2	3	2
Phenylalanine	4	8	4	4	7
Lysine	10	18	9	9	8
Histidine	4	8	4	4	4
Arginine	3	6	3	3	7
Tryptophan	3	4	2	3	3
Total	117	230	115	116	118

[a] From Angeletti (1970). [b] From Angeletti et al. (1973a).

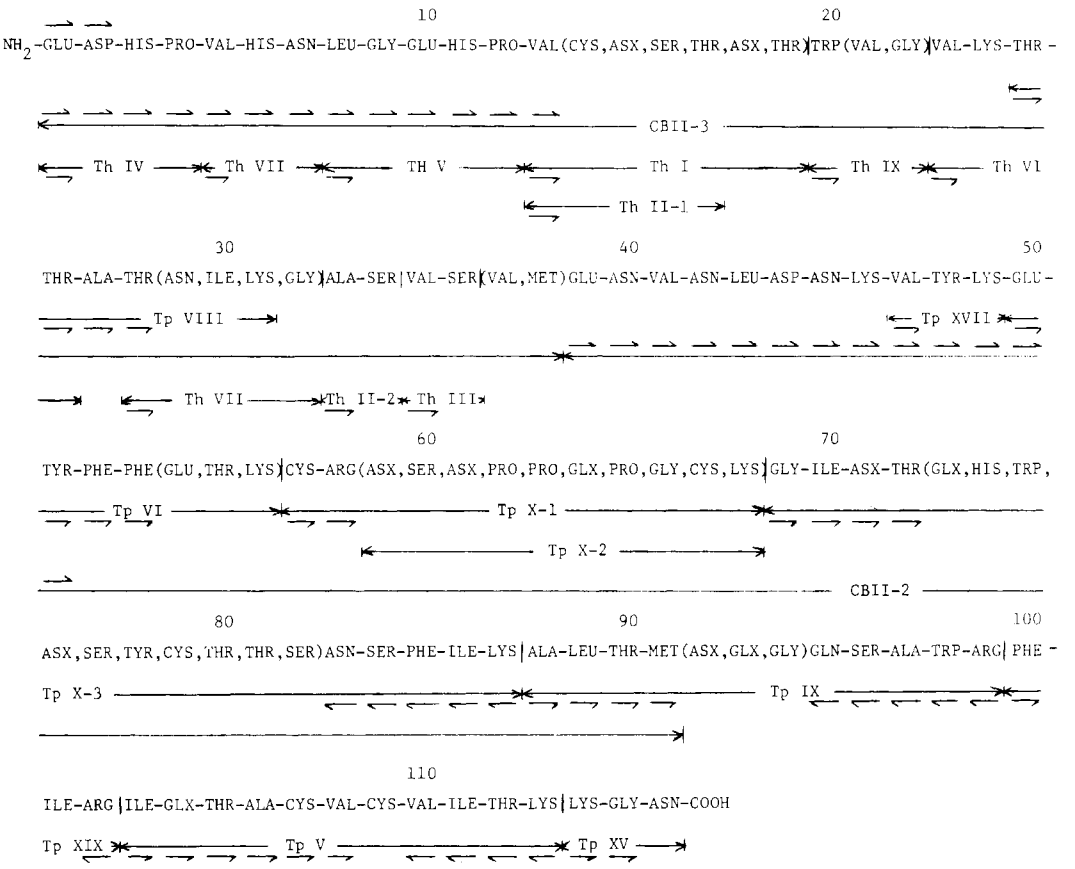

FIGURE 8: The tentative amino acid sequence of *Naja naja* nerve growth factor. The various peptides are indicated by double-headed arrows. Residues identified by manual Edman degradation (—), Sequencer analysis (—), and carboxypeptidase A and/or B hydrolysis (—) are so indicated. Abbreviations used are: CB, cyanogen bromide; Tp, tryptic; Th, thermolytic. Vertical slashes indicate residues thought to be contiguous by comparison to the mouse NGF sequence, but for which rigorous internal overlaps are lacking.

Naja naja (6.75) native proteins (Bocchini and Angeletti, 1969; Angeletti, 1970).

From the amino acid composition of the *Naja naja* NGF, 14 tryptic peptides containing 6 residues of radioactively labeled half-cystine were expected on the basis of the minimum molecular weight of 14000. Twice that number would be expected if the dimer were composed of nonidentical subunits. In fact, ten soluble tryptic peptides containing five radioactively labeled half-cystinyl residues were isolated in significant yield. These data confirm that the subunit structure of *Naja naja* NGF is similar to that of mouse NGF, that is, a native protein of about 28000 molecular weight consisting of two identical subunits.

Figure 8 summarizes the data obtained from the sequence analysis of the tryptic peptides and cyanogen bromide fragments of *Naja naja* NGF. Three cyanogen bromide fragments were observed, two of which contained 54 and 37 residues, indicating that neither methionine corresponded to the methionine at position 9 in the mouse NGF sequence. Although the third fragment was never purified to homogeneity, it contained no trace of homoserine and homoserine lactone, and was thus judged to be derived from the carboxyl terminus of the molecule. The amino-terminal analysis of the whole protein in the protein Sequencer was successfully pursued on only two residues, which coincided with the amino-terminal sequence of CB II-3, Glu-Asp. From the thermolytic peptides derived from this fragment, peptides Th IV, Th VIII, and Th V had amino acid compositions and amino-terminal residues consistent with the portions of this fragment established by the protein Sequencer. Homology with the mouse protein (vide infra) strongly supports the assignment of half-cystine as the next residue which, in turn, allows the placement of peptides Th I and Th II-1 as next in the sequence. The following two peptides, Th IX and Th VI, were assigned by homology of amino acid composition and amino-terminal analysis to the mouse structure. Similarly, Tp VIII and Th VII provided identifications through residue 32. The two dipeptides remaining, Th II-2 and Th III, were placed as residues 33 to 36 by comparison to the mouse NGF sequence and the remaining amino acid composition of fragment CB II-3. No homoserine-containing peptide was recovered.

The first 13 residues of CB II-2 were established unequivocally by the use of the protein Sequencer. A tryptic peptide corresponding to the amino-terminal portion of this fragment was not found. Peptides Tp XVII and Tp III permitted the structure to be extended another two residues. Since attempts to obtain thermolytic peptides from this fragment in sufficient yield were unsuccessful, only sequential homology to the mouse protein was used to establish the following residues. Tp X-1 and Tp X-2 which differed by only two residues corresponded to the two contiguous half-cystine-containing peptides in the mouse NGF sequence. These two peptides proved to be refractive to Edman degradation, however, and the internal sequence is arbitrarily assigned, based as much as possible on residue to residue homology with the mouse NGF sequence. The amino-terminal sequence of Tp X-3 established its homology to the adjacent peptide of the mouse sequence; however, the carboxy-terminal sequence corresponded to the mouse tryptic

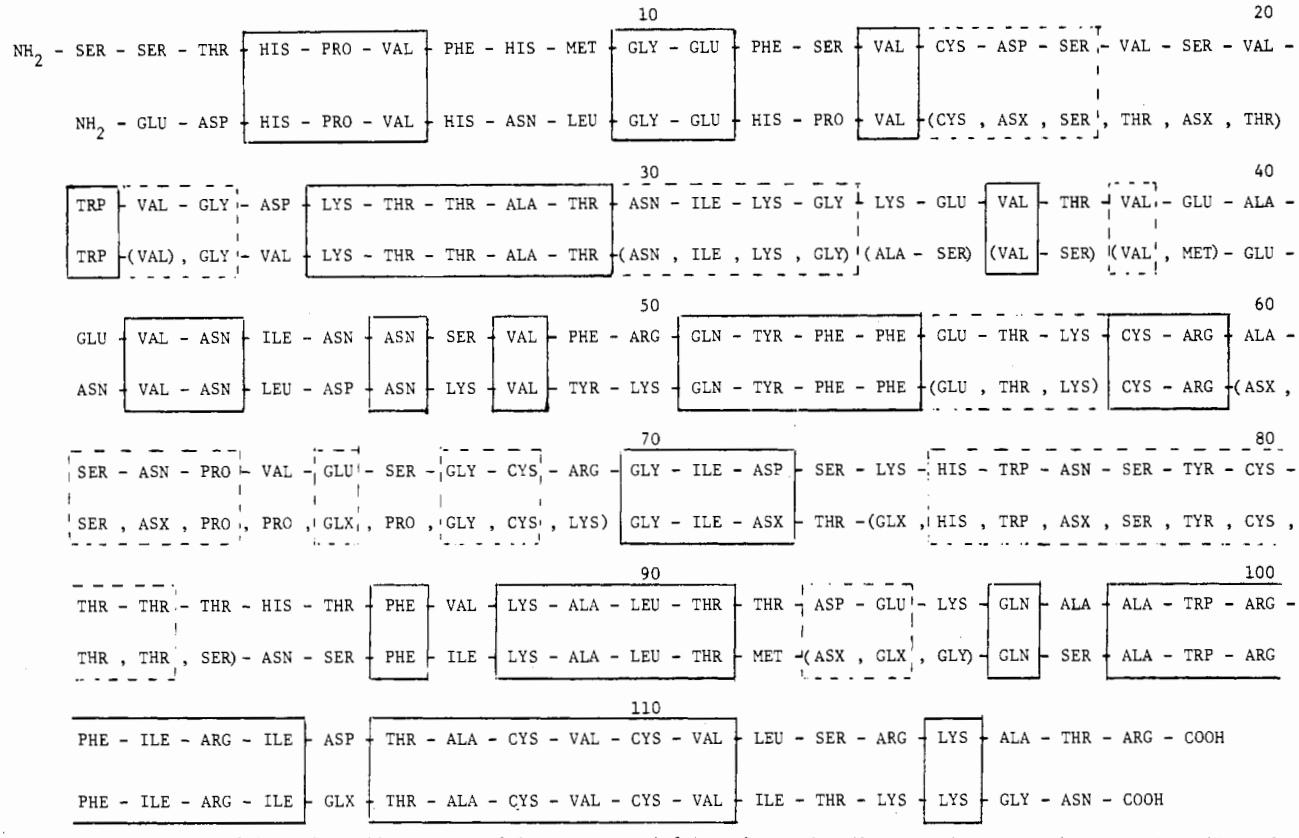

FIGURE 9: Comparison of the amino acid sequences of the nerve growth factors from cobra (*Naja naja*) venom and mouse submaxillary glands. Solid boxes enclose rigorously positioned identical residues. Dashed boxes enclose identities positioned by homology.

peptide next in sequence, suggesting the replacement of a lysine residue. Peptide Tp IX encompassed a region in the mouse sequence which contained two tryptic peptides. Furthermore, Edman degradation firmly positioned the second methionine at residue 91. The remainder of the sequence was established by the peptides Tp IX, Tp XIX, and Tp V which were identical or very similar in structure to those isolated from the mouse protein. Tp XV had amino-terminal lysine and carboxy-terminal asparagine and was thus logically assigned to the carboxy terminus of the polypeptide chain.

By positioning cyanogen bromide fragments and tryptic peptides in this manner the maximum amount of homology to the mouse NGF structure can be obtained. In particular, the assignment of CB II-3 to the amino terminus of the polypeptide chain allows this half-cystine-poor fragment to align to the half-cystine-poor region of mouse NGF. Conversely, B II-2 is half-cystine-rich and aligns well with the middle region of the mouse NGF sequence. Furthermore, the methionine-containing tryptic peptide Tp IX positions CB II-2 firmly to this region. Table VIII shows that the amino acid compositions of CB II-2 and CB II-3, calculated from data and from the sequence, are within experimental error. The amide assignments of residues 29, 54, and 116 were determined by high-voltage paper electrophoresis at pH 6.5. Residues 83 and 95 were established by carboxypeptidase A digestion. The other clearly identified amide assignments were established by thin-layer chromatography of the Pth derivatives obtained from the protein Sequencer.

It should be stressed that construction of a complete tentative sequence for *Naja naja* NGF relies heavily on the high degree of similarity with mouse NGF. The regions of the *Naja naja* chain where sequence data for tryptic peptides and cyanogen bromide fragments have been obtained

Table VIII: Experimental and Theoretical Amino Acid Composition of Cyanogen Bromide Fragments of *Naja naja* NGF.

	CB II-2		CB II-3	
	Residues/mol		Residues/mol	
Amino Acid	Exptl	Sequence	Exptl	Sequence
Aspartic acid	9	9	5	5
Threonine	5	5	5	5
Serine	4	4	3	3
Glutamic acid	6	5	2	2
Proline	3	3	2	2
Glycine	2	2	2	3
Alanine	1	1	2	2
Half-cystine	3	3	1	1
Valine	2	2	5	6
(Methionine)[a]	(1)	1	(1)	1
Isoleucine	1	2	1	1
Leucine	2	2	2	1
Tyrosine	3	3		
Phenylalanine	3	3		
Lysine	6	5	2	2
Histidine	1	1	3	3
Arginine	1	1		
Tryptophan[b]	(1)	1	(1)	1
Total	54	53	37	38

[a] As homoserine. [b] Estimated by the spectrophotometric method of Edelhoch (1967).

comprise approximately 60% of the molecule and, in these regions, the number of identities is greater than 60%. Thus, the construction of the complete tentative sequence relying on homology with the mouse protein is justified. Had the degree of similarity between *Naja naja* and mouse NGF been less than the substantial degree observed in the struc-

tured regions, the data which it was possible to collect using the small amount of material available would not have been sufficient to allow the construction of the tentative sequence.

This tentative partial primary structure of *Naja naja* NGF is shown in Figure 9 aligned with the mouse NGF sequence. Of the 73 residues which have been independently sequenced (Figure 8), 44 residues align with identical residues in the mouse sequence, yielding 61% (44/73) identity in these regions. The remaining 43 residues of the venom NGF sequence are accounted for in peptides which have been positioned in the structure by aligning the residues to yield the maximum degree of identity with the mouse sequence. The identities resulting from this alignment of the unstructured peptides to give a best fit are enclosed in dashed boxes in Figure 9. These 30 additional identities occurring in the 43 unstructured residues yield an extent of identity of 70%, a value not greatly different than the 61% observed in the structured regions of the sequence. The total number of possible identities shown in Figure 9 is 74 of 116 or 64%. It should be noted that, even if only those identities which occur in directly sequenced portions of the venom NGF peptide chain are considered, the minimum identity is still 44 residues or 38% of the total polypeptide chain. This degree of identity is itself unambiguously indicative of genetic relatedness. However, in view of the 60% identity observed in the structured region of the venom NGF peptide chain which account for nearly two-thirds of the molecule, it is likely that the true extent of similarity is closer to the calculated value of 64% identical residues.

The alignment of the *Naja naja* NGF and mouse NGF sequences shown in Figure 9 places all six half-cystinyl residues of each protein in identical positions. Considering the high degree of similarity between the proteins, it is very probable that the six half-cystinyl residues are arranged in an identical disulfide configuration in the two molecules. The distribution of identical residues in the sequence alignment (Figure 9) appears to be quite uniform and thus no particular segments of the comparison stand out as being more strongly conserved, and thus essential for function, than others. With few exceptions nonidentical residues may be described as replacements conservative of chemical function.

The partial sequence data obtained for *Naja naja* NGF have provided a conclusive answer to the primary question which motivated this study. It is clear that cobra venom NGF is structurally similar to mouse NGF to a degree obviously indicative of the close genetic relatedness found among members of functionally related families of proteins. Furthermore, it is now possible to examine in better perspective the evolutionary events which relate this family of NGF proteins to the family of proinsulins (Frazier et al., 1972). The 64% identity of mammalian and reptilian NGF's compares with the 70% identity of fish and mammalian insulins (Dayhoff, 1969). Comparison of *Naja naja* NGF with the family of proinsulins does not indicate a greater degree of relatedness than that previously observed between mouse NGF and the proinsulins (Frazier et al., 1972). Therefore, it would appear that the divergence of the NGF and the insulin families occurred very early in vertebrate evolution and that selective pressure has maintained the structure of NGF as rigidly as that of the proinsulins.

A comparison of other functional and structural properties of *Naja naja* and mouse NGF is described in the following paper (Server et al., 1976).

Acknowledgments

The authors wish to thank Ms. Linda F. Boyd for performing the bioassays. A portion of the crude venom was provided for R.A.H.A. by the Laboratorio di Biologia Cellulare (Consiglio Nazionale delle Ricerche) in Rome, Italy, at the inception of these studies.

References

Angeletti, R. H. (1970), *Proc. Natl. Acad. Sci. U.S.A. 65,* 668.

Angeletti, R. H. (1971), *Brain Res. 25,* 424.

Angeletti, R. H., Bradshaw, R. A., and Wade, R. D. (1971), *Biochemistry 10,* 463.

Angeletti, R. H., Hermodson, M. A., and Bradshaw, R. A. (1973a), *Biochemistry 12,* 100.

Angeletti, R. H., Mercanti, D., and Bradshaw, R. A. (1973b), *Biochemistry 12,* 90.

Bocchini, V., and Angeletti, P. U. (1969), *Proc. Natl. Acad. Sci. U.S.A. 64,* 787–794.

Boyd, L. F., Bradshaw, R. A., Frazier, W. A., Hogue-Angeletti, R. A., Jeng, I., Pulliam, M. W., and Szutowicz, A. (1974), *Life Sci. 15,* 1381.

Bradshaw, R. A., Babin, D. R., Nomoto, M., Srinivasin, N. G., Ericsson, L. H., Walsh, K. A., and Neurath, H. (1969), *Biochemistry 8,* 3859.

Davies, B. J. (1964), *Ann. N.Y. Acad. Sci. 161,* 404.

Dayhoff, M. O. (1969), Atlas of Protein Sequence and Structure, Vol. 4, Silver Spring, Md., National Biomedical Research Foundation.

Edelhoch, H. (1967), *Biochemistry 6,* 1948.

Edman, P. (1956), *Acta Chem. Scand. 10,* 761.

Frazier, W. A., Boyd, L. F., and Bradshaw, R. A. (1974a), *J. Biol. Chem. 249,* 5513.

Frazier, W. A., Boyd, L. F., Pulliam, M. W., Szutowicz, A., and Bradshaw, R. A. (1974b), *J. Biol. Chem. 249,* 5918.

Frazier, W. A., Hogue-Angeletti, R., and Bradshaw, R. A. (1972), *Science 176,* 482.

Frazier, W. A., Hogue-Angeletti, R. A., Sherman, R., and Bradshaw, R. A. (1973), *Biochemistry 12,* 3281.

Gray, W. R., and Smith, J. F. (1970), *Anal. Biochem. 33,* 36.

Gross, E., and Witkop, B. (1961), *J. Am. Chem. Soc. 83,* 1510.

Hartley, B. S. (1970), *Biochem. J. 119,* 805.

Hill, R. L., and Delaney, R. (1967), *Methods Enzymol. 11,* 339.

Jacobs, J. W., and Niall, H. D. (1975), *J. Biol. Chem. 250,* 3629.

Konigsberg, W. J., and Hill, R. J. (1962), *J. Biol. Chem. 237,* 2547.

Levi-Montalcini, R., and Angeletti, P. U. (1968), *Physiol. Rev. 48,* 534.

Levi-Montalcini, R., and Cohen, S. (1956), *Proc. Natl. Acad. Sci. U.S.A. 42,* 571.

Moore, S. (1963), *J. Biol. Chem. 238,* 235.

Ornstein, L. (1964), *Ann. N.Y. Acad. Sci. 161,* 321.

Server, A. C., Herrup, K., Shooter, E. M., Hogue-Angeletti, R. A., Frazier, W. A., and Bradshaw, R. A. (1976), *Biochemistry,* following paper in this issue.

Weber, K., and Osborn, M. (1969), *J. Biol. Chem. 244,* 4406.

Woods, K. R., and Wang, K. T. (1967), *Biochim. Biophys. Acta 133,* 369.

Yphantis, D. A. (1964), *Biochemistry 3,* 297.