Proc. Nat. Acad. Sci. USA
Vol. 68, No. 10, pp. 2417–2420, October 1971

Nerve Growth Factor from Mouse Submaxillary Gland: Amino Acid Sequence

(2.5S protein/two subunits/disulfide bonds/enzymatic digestion)

RUTH HOGUE ANGELETTI* AND RALPH A. BRADSHAW†

* Department of Biology, Washington University, St. Louis, Missouri 63130; and † Department of Biological Chemistry, Washington University School of Medicine, St. Louis, Missouri 63110

Communicated by Rita Levi-Montalcini, July 29, 1971

ABSTRACT **The complete amino-acid sequence of the 2.5S nerve growth factor from male-mouse submaxillary glands has been determined. The unambiguous alignment of peptides derived from tryptic, chymotryptic, thermolytic, and peptic digestion of *S*-carboxymethyl-, *S*-aminoethyl-, and native growth factor indicates that the primary subunit is composed of 118 amino acids, with amino-terminal serine and carboxyl-terminal arginine. The molecular weight of this subunit, calculated from the primary sequence, is 13,259. Thus, the native protein, which is composed of two of the subunits, has a molecular weight of 26,518. These values, as well as the final amino-acid composition, are in excellent agreement with those determined by direct measurement with undigested growth factor. The alignment of the three disulfide bonds, determined from a combination of peptic and thermolytic digestions, is I-IV, II-V, and III-VI. The latter two pairs are located in a closed loop of 14 amino acids, by virtue of the fact that half-cystinyl residues V and VI are separated by only a single residue in the linear sequence. Assignment of the side-chain amides showed that 7 of 11 aspartic acid residues and 2 of 8 glutamic acid residues are present as amides. This distribution of charged residues is entirely consistent with the observed isoelectric point of 9.3.**

The ability of a specific protein to enhance the growth of sympathetic ganglia *in vitro* and *in vivo*, and of sensory ganglia *in vitro*, was first demonstrated by Levi-Montalcini (1). This compound, designated nerve growth factor (NGF) (2), has been demonstrated in low concentrations in a wide variety of tissues, and in much higher concentrations in certain snake venoms and in male-mouse submaxillary glands (3, 4). Studies with purified mouse-NGF preparations, which have detailed the metabolic and cellular responses to NGF (5), indicate the importance of this factor as a regulator of cell growth and differentiative processes in neural tissue (6). However, chemical and structural characterization of mouse NGF, and the relation of its structure to its mechanism of action, have been less complete. Initial studies indicated that the growth-promoting activity is associated with a protein of about 30,000 molecular weight, with a sedimentation constant of 2.5 S (7). The first amino-acid analyses reported indicated that the protein contained 265 amino acids and had an apparent isoelectric point of 9.3 (8). An additional degree of complexity concerning the relation of structural and functional properties of NGF was introduced with the report that the factor is associated with at least two other proteins (denoted α and γ) when it is first released from the tissue (9). Although all of the nerve growth-promoting activity of this complex (designated 7S NGF) is associated with the 2.5S subunit (denoted β), it has been reported that additional activity can be regenerated by recombination of the β subunit with previously removed α and γ subunits (10). In contrast, preparations of 2.5S NGF, as prepared by the method of Bocchini and Angeletti (7), appear to account for all of the NGF activity initially present in tissue homogenates.

In order to clarify these problems and to provide a basis for determining the mechanism of action of NGF in molecular terms, we began studies to elucidate the structure of this protein. By means of sedimentation analyses, amino end-group determinations, and fractionation of the soluble tryptic peptides, we ascertained that 2.5S NGF has a dimeric structure composed of two very similar or identical subunits, of molecular weight about 14,500 (11). It was further established that each polypeptide chain possessed three disulfide bonds, and that the subunits were associated in the native state by noncovalent forces. These observations have now been extended to include the complete amino-acid sequence of the NGF subunit. These results indicate that NGF is, indeed, composed of two identical subunits, although slightly shorter chains, which have been generated by proteolysis, have also been isolated in the form of mixed dimers.

DETERMINATION OF THE AMINO-ACID SEQUENCE

All of the sequence analyses were performed on 2.5S NGF, as prepared by the method of Bocchini and Angeletti (7). The final structure, as shown schematically in Fig. 1, was deduced from five separate enzymatic digestions of the protein. A summary of these experiments is listed in Table 1. About 30–50 mg of native or modified NGF was used in each digest. The soluble peptides from each digest were purified on columns (0.9 × 20 cm) of the indicated resins, with pyridine acetate gradients as described (12). Further purifications, analyses for purity, and determination of amino-acid composition were by established techniques (13). Sequences of homogeneous peptides were determined by substractive Edman degradation (14), by direct identification of methylthiohydantoins by gas–liquid chromatography (15), and by the dansylation method (16). Carboxyl-terminal residues were determined by hydrolysis with carboxypeptidases A and/or B (14). Side-chain amides of glutaminyl and asparaginyl residues were determined from electrophoretic mobilities at pH 6.5, and from direct identification of thiohydantoins after Edman degradation.

Abbreviation: NGF, nerve growth factor.

* Present address: Laboratorio di Biologia Cellulare, Rome, Italy.

† To whom correspondence may be addressed.

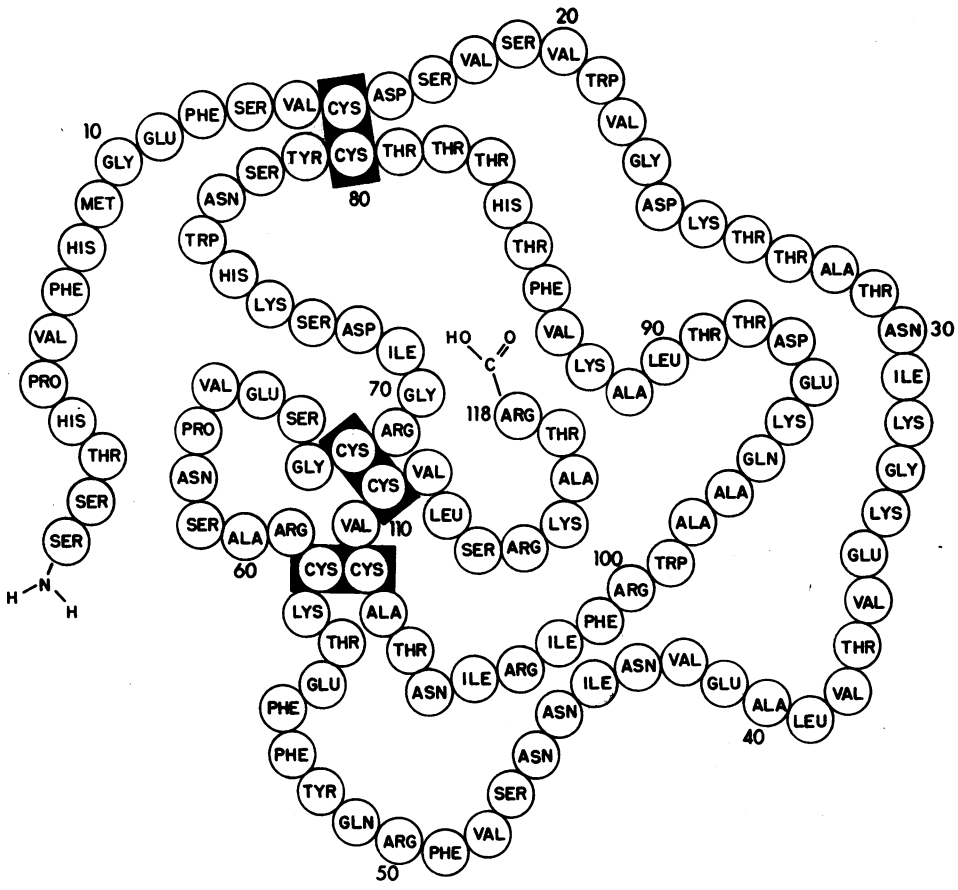

FIG. 1. Schematic representation of the amino-acid sequence of the primary subunit of 2.5S NGF from mouse submaxillary glands.

In addition to the enzymic digests described, sequence information was also obtained by analysis of the protein by the amino-acid sequenator. These experiments unambiguously identified the initial 33 residues of the polypeptide chain and were in complete agreement with the data obtained from the isolated peptides‡.

The carboxyl-terminal sequence of the whole protein was determined by digestion with carboxypeptidase B, followed by carboxypeptidase A. *S*-carboxymethyl NGF was used as substrate.

DETERMINATION OF THE DISULFIDE BONDS

The correct alignment of the half-cystinyl residues was determined from combined peptic and thermolytic digests of native NGF. After fractionation of the peptic digest of NGF on SE-Sephadex C-25 at 55°C (17), it was ascertained by cyanide–nitroprusside staining (18) of electrophoretograms that all three disulfide bonds were contained in a single peptide. Consequently, the intact peptic peptide was digested with thermolysin at pH 7 to minimize disulfide interchange. Fractionation of this digest on columns of SE-Sephadex, resulted in the isolation of three separate disulfide peptides, as well as many nondisulfide peptides. Each disulfide peptide was oxidized with performic acid (19), and the resulting mixture was fractionated on columns of Dowex 50 × 8 and by paper electrophoresis. Each oxidized peptide was identified by its amino-acid composition.

‡ Hermodsen, M., R. H. Angeletti, W. A. Frazier, and R. A. Bradshaw, manuscript in preparation.

DISCUSSION

Examination of the complete amino-acid sequence depicted in Fig. 1 reveals that the primary subunit of NGF is composed of 118 amino acids. Analyses performed in the protein sequenator have also revealed that a shorter chain, identical to the primary subunit except for the absence of the first 8 amino acids, can also be isolated. The occurrence of this chain appears to be confined to homogeneous dimers containing one long and one short chain, and is probably dependent upon the method of preparation of NGF. These results, to be described in detail‡, may also provide an explanation for some of the differences observed between 2.5S NGF and the β-subunit of 7S NGF.

The amino-acid composition and molecular weight calculated from the primary structure are in good agreement with the values obtained from whole NGF (11). This calculated molecular weight for the subunit, 13,259, may be compared with the value of 14,530 ± 1432 determined for *S*-carboxymethyl NGF in 6 M guanidine·HCl by sedimentation equilibrium methods. The amino-acid composition calculated from acid hydrolysates, when compared to that determined from the sequence analysis, was one residue too high in lysine, aspartic acid, alanine, valine, and leucine, and two residues too high in glutamic acid and glycine. However, the calculated composition was based upon a molecular weight of 14,500. Recalculation of the data obtained from acid hydrolysates on the basis of the exact molecular weight largely corrects these discrepancies.

TABLE 1. *Summary of the enzymatic digests for the sequence analysis of mouse 2.5S NGF*

Derivative	Proteolytic enzyme	pH	Temperature, °C	Initial fractionation medium	Remarks
[^{14}C]*S*-carboxymethyl NGF	Trypsin	8.8	40	Dowex 50 × 8	Insoluble material at pH 2.5 removed before application to column
S-aminoethyl NGF	Trypsin	8.8	40	Dowex 50 × 8	Insoluble material at pH 2.5 removed before application to column
[^{14}C]*S*-carboxymethyl NGF	Chymotrypsin	8.5	40	Dowex 50 × 8	Insoluble material at pH 2.5 removed before application to column
S-carboxymethyl NGF	Thermolysin	8.0	40	Dowex 50 × 8	No insoluble material
Native NGF	Pepsin	2.0	25	SE-Sephadex C-25	Insoluble material at pH 2.5 removed before application to column

Pyridine acetate was used to elute the columns.

Several features of the primary structure should be noted. The amino-terminal portion of the molecule is considerably less basic than the carboxyl-terminal region. In fact, only three lysines and one arginine are found in the first 50 amino acids. This distribution is reflected in the fact that the tryptic-insoluble core from *S*-carboxymethyl NGF, which is composed of two peptides covering residues 1–25 and 35–50, accounts for over 80% of these 50 residues§. The chymotryptic core is also derived mainly from this region (manuscript in preparation).

Conversely, the carboxyl-terminal portion of the molecule is more basic in character and, as judged by the alignment of the disulfide bonds, exists in a more rigidly confined conformation. The most striking feature revealed by the disulfide structure is the presence of a 14-residue loop that is formed by closing the disulfide bonds between half-cystines II-V and III-VI. This structure, because of the Cys-Val-Cys sequence at positions 108–110, presented a technical problem in that pepsin did not cleave either of the peptide bonds joining these three residues. It is of interest to note that thermolysin, which has an excellent specificity for valine (20), was capable of hydrolyzing the Cys-Val bond that joined residues 108–109, even with both disulfide bonds intact. This observation suggests that the active site of thermolysin may be more spacious than that associated with other proteolytic enzymes, which are normally inhibited by the presence of bulky groups on neighboring residues.

The assignment of the amide side-chains of the asparaginyl and glutaminyl residues revealed that 7 of 11 aspartate and 2 of the 8 glutamate residues are present in the amide form. Thus, only 10 of the 19 potential acidic residues are found in the free-acid form. This distribution of charge is in excellent agreement with the observed pI of 9.3, determined by isoelectric focussing (8), and accounts for the basic behavior of this protein on negatively charged ion-exchange resins (7).

The identification of the carboxyl-terminal residue as arginine is of interest with regard to the observations of Taylor *et al.* (21) that epidermal growth factor (EGF), a polypeptide of molecular weight 6000 also isolated from mouse submaxillary glands, has carboxyl-terminal arginine as well. It has been proposed that this factor may be released from a larger protein by the action of an arginine-specific protease present in the gland. Although there is at present no evidence to suggest that NGF is released from a larger protein by the action of such an enzyme, such possibilities are currently under investigation.

A comparison of the sequence of NGF with that of several other proteins has not revealed any obvious structural relationships of the type associated with homologous proteins. In particular, lysozyme, α-lactalbumin, and ribonuclease A, which possess similar numbers of residues, have been examined without success for similarities maintained through evolution from a common precursor. While this limited analysis does not preclude the identification of other proteins whose structures have already been determined as sharing a common precursor with NGF, it does not allow, at present, any further insight into the biological activity of this factor.

The completion of the primary-structure analysis of NGF provides the opportunity for many further experiments that will help to clarify the role of this factor in physiological processes. It may be hoped that when all of the structural features of NGF have been defined that it will be possible to completely describe, in molecular terms, the growth and differentiative processes of sympathetic nerve tissue.

We express our appreciation to Mr. Delio Mercanti for his excellent technical assistance throughout this work, and to Mr. Charles Yates and Mr. Robert Sherman for their assistance with the amino-acid analyses. We also express our gratitude to Dr. Rita Levi-Montalcini for her financial and intellectual support of this project. This investigation has been supported by grants from the National Institutes of Health, AM-13362 and NS-;03777. RAB is a Research Career Development Awardee of the NIH (AM-23968).

1. Levi-Montalcini, R., *Ann. N.Y. Acad. Sci.*, **55**, 330 (1952).
2. Cohen, S., R. Levi-Montalcini, and V. Hamburger, *Proc. Nat. Acad. Sci. USA*, **40**, 1014 (1954).
3. Levi-Montalcini, R., and S. Cohen, *Proc. Nat. Acad. Sci. USA*, **42**, 695 (1956).
4. Levi-Montalcini, R., and S. Cohen, *Ann. N.Y. Acad. Sci.*, **85**, 324 (1960).

§ The tryptic peptides of 2.5S NGF and the β-subunit of 7S NGF have been isolated and characterized independently by us and by D. Straus, W. DeJong, and E. Shooter, Stanford University. These results will be published as a joint communication.

5. Levi-Montalcini, R., and P. U. Angeletti, *Physiol. Rev.*, **48**, 534 (1968).
6. Levi-Montalcini, R., and P. U. Angeletti, in *Ciba Found. Symp. on Growth of the Nervous System*, eds. G. E. W. Wolstenholme and M. O'Connor (Churchill, London, 1968), p. 126.
7. Bocchini, V., and P. U. Angeletti, *Proc. Nat. Acad. Sci. USA*, **64**, 787 (1969).
8. Bocchini, V., *Eur. J. Biochem.*, **15**, 127 (1970).
9. Varon, S., J. Nomura, and E. M. Shooter, *Biochemistry*, **6**, 2202 (1967).
10. Varon, S., J. Nomura, and E. M. Shooter, *Proc. Nat. Acad. Sci. USA*, **57**, 1782 (1967).
11. Angeletti, R. H., R. A. Bradshaw, and R. D. Wade, *Biochemistry*, **10**, 463 (1971).
12. Bradshaw, R. A., G. W. Robinson, G. M. Hass, and R. L. Hill, *J. Biol. Chem.*, **244**, 1755 (1969).
13. Bradshaw, R. A., W. H. Garner, and F. R. N. Gurd, *J. Biol. Chem.*, **244**, 2149 (1969).
14. Bradshaw, R. A., D. R. Babin, M. Nomoto, N. G. Srinivasin, L. H. Ericsson, K. A. Walsh, and H. Neurath, *Biochemistry*, **8**, 3859 (1969).
15. Waterfield, M., and E. Haber, *Biochemistry*, **9**, 832 (1970).
16. Hartley, B. S., *Biochem. J.*, **119**, 805 (1970).
17. Walsh, K. A., R. M. McDonald, and R. A. Bradshaw, *Anal. Biochem.*, **35**, 193 (1970).
18. Toennies, G., and J. J. Kolb, *Anal. Chem.*, **23**, 823 (1951).
19. Hirs, C. H. W., *Methods Enzymol.*, **XI**, 197 (1967).
20. Bradshaw, R. A., *Biochemistry*, **8**, 3871 (1969).
21. Taylor, J. M., S. Cohen, and W. M. Mitchell, *Proc. Nat. Acad. Sci. USA*, **67**, 164 (1970).