The Covalent and Three-Dimensional Structure of Concanavalin A

II. AMINO ACID SEQUENCE OF CYANOGEN BROMIDE FRAGMENT F₁

(Received for publication, June 10, 1974)

BRUCE A. CUNNINGHAM, JOHN L. WANG, MYRON J. WAXDAL,† AND GERALD M. EDELMAN

From The Rockefeller University, New York, New York 10021

SUMMARY

The amino acid sequence of the COOH-terminal CNBr fragment, F₁ (residues 130 to 237), of concanavalin A has been established, completing the determination of the covalent structure of this lectin. Analysis of the chemical sequence showed that the distribution of charged residues is generally more dense in the NH₁-terminal half of the polypeptide chain than in the COOH-terminal portion and that in the latter region there is a linear stretch composed of many hydrophobic residues. Correlation with x-ray crystallographic results showed that the distribution of charged residues is generally more dense in the NH₁-terminal half of the polypeptide chain than in the COOH-terminal portion and that in the latter region there is a linear stretch composed of many hydrophobic residues. In conjunction with the three-dimensional structure of this lectin. Analysis of the chemical sequence provides new data for analysis of variables involved in predicting the three-dimensional folding of proteins from the primary structure. The sequence of concanavalin A is the first determined for a lectin and it serves as a reference structure for comparisons with other lectins.

Studies on the structure and activities of the lectin Con A have been pursued in a number of laboratories in order to understand the wide variety of biological and ligand-binding activities exhibited by this protein (see Ref. 1). The over-all structure of Con A at physiological pH has been shown to be a tetramer of identical subunits (2-4), each with a single saccharide binding site (5, 6).

In addition to studies to characterize the molecular structure of Con A, other experiments have attempted chemical modification of the polypeptide chain in an effort to alter the activities of the protein. These studies include the labeling of the metal and saccharide binding sites (7, 8), the dissociation of the subunits of the tetramer (9, 10), and various other derivatizations using group-specific reagents (11, 12). The effects of proteolytic cleavages of the polypeptide chain (13, 14) have also been examined. Such studies can now be refined further to correlate the modification of specific amino acid residues with corresponding alterations in the activities and to interpret these changes in terms of the complete chemical sequence and three-dimensional structure of the Con A molecule.

In the preceding paper of this series (1), we presented the amino acid sequence of the first 129 residues of the Con A polypeptide chain. We now report the complete covalent structure of the protein. This sequence represents the first complete chemical structure of a lectin and thus provides a reference for comparison with other mitogenic and non-mitogenic lectins. Correlation of the amino acid sequence with the results of x-ray crystallographic studies are reported in the following papers (15, 16).

MATERIALS AND METHODS

The isolation of the intact subunit of Con A (13) and the preparation of CNBr Fragment F₁ (17) have been described. Fragment F₁ was separated from Fragments F₁ and F₂ by gel filtration on Sephadex G-75 in 5% formic acid and purified by repeated gel filtration on the same column. Peptide T₁₅₁₆ was succinylated by suspension in saturated sodium acetate (9) and addition of a 100-fold molar excess of succinic anhydride. After 1 hour at room temperature, an equal volume of 2 M propionic acid was added and the reaction mixture was stirred overnight. The resulting suspension was centrifuged and the precipitate was dissolved in AmProl and separated from any unreacted reagent by gel filtration on Sephadex G-50 in AmProl.

Enzymatic digests were performed as described previously (1). Partial acid hydrolysis of peptides was carried out in 0.03 M HCl, 0.1 M HCl, or 0.36 M acetic acid at 110° for the indicated times in tubes sealed in vacuo. Tubes were then opened and the contents were dried by evaporation in vacuo over NaOH or by lyophilization prior to fractionation.

Procedures used for the isolation of peptides by gel filtration and ion exchange chromatography have been reported (18-20). High voltage paper electrophoresis was carried out at pH 4.5 and 6.5 (21). Descending paper chromatography of peptides was performed in n-butanol-acetic acid-water-pyridine, 15:3:12 (v/v/v) (23). Procedures employed for amino acid analysis (24), the determination of the amino acid sequence of peptides (18, 25), and the detection of tryptophan in peptides...
(26) are identical with those described in the previous paper of this series (1). Direct sequence analysis of CNBr Fragment F, was carried out with a Beckman model 890C Sequencer (Beckman Instruments) using the Quadrol double cleavage program (Beckman model 890C Sequencer Instruction Manual). Phenylthiohydantoinas were identified by thin layer chromatography on Chromagram silica gel sheets with fluorescent indicator (Eastman Kodak Co.) using Solvent Systems V and IV described by Jeppson and Sjöquist (27), and benzene-acetic acid (9:1, v/v) (20).

RESULTS

Amino Acid Sequence of CNBr Fragment F,—Previous studies (1) have established the amino acid sequence of CNBr Fragments F, and F, representing residues 1 to 129 of Con A. The remainder of the amino acid sequence was established by the analysis of peptides isolated from enzymatic digests of Fragment F, which contains the rest of the peptide chain (residues 130 to 257) with an NH2-terminal sequence of Phe-Ass-Glu-Phe-Ser-Lys and a COOH-terminal sequence of Asx-Ala-Asn (17). The numbering system used here is identical and continuous with that used in the preceding paper (1). Few peptides were required to establish the order of the tryptic peptides and each of the segments that were difficult to analyze is contained within the regions represented by a tryptic peptide. The determination of the amino acid sequence of Fragment F, therefore, is discussed in terms of the tryptic peptides as outlined in Fig. 1.

Tryptic Peptides of F,—The amino acid compositions of the tryptic peptides obtained from Fragment F, are tabulated in Table I. Fragment F, (10 μmol) was digested with trypsin for 4 hours, the solution lyophilized, and the residue extracted with 0.02 M Tris, pH 8.0, to give a supernatant and a precipitate fraction. The precipitate was used to obtain peptides T15,16a and T15,16 as described below. The soluble fraction of the tryptic digest was chromatographed on DEAE-cellulose (Fig. 2). Material in Fraction A was further fractionated by gel filtration on Sephadex G-25 in AmProP to give peptides T11b and T14. Material in Fractions B, C, and F (Fig. 2) yielded T12, T13, and T17, respectively, without further purification.

Material from the tryptic digest that was insoluble in 0.02 M Tris, pH 8.0, contained peptides represented by the designation T15,16 (Fig. 1). Neither peptide T15 nor T16 could be isolated as such; instead, peptides T15,16, T15,16a, and T16b were obtained. Although peptides T15,16 and T15,16a presumably resulted from the failure of trypsin to cleave the peptide bond between Lys 200 and Ser 201, T15,16 differed from T15, 16a. The latter peptide apparently resulted from chymotryptic cleavage of the peptide bond between Phe 212 and Phe 213. Peptides T15,16a and T16b were obtained from the same digests, although T15,16a was frequently contaminated with T15,16. Peptide T16b was soluble in 0.02 M Tris, pH 8.0, and was also obtained from the soluble fraction of a tryptic digest of the intact subunit of Con A by gel filtration on Sephadex G-50 in AmProP followed by high voltage electrophoresis at pH 4.7 and 1.9. When rigorous efforts were made to reduce the extent of chymotryptic cleavage (using trypsin with minimal chymoactivity and limiting digestion to 4 hours) peptide T15, 16 predominated and no T16b was obtained.

Peptides T15,16a and T15,16 were obtained from material insoluble in 0.02 M Tris, pH 8.0, by gel filtration on Sephadex...
TABLE I

Amino acid composition of trypsic peptides from CNBr fragment F₃

| Tube | T11b | T12 | T13 | T14 | T15,16a | T16b | T15,16 | T17 |
|------|------|-----|-----|-----|---------|------|--------|-----|
| Lys  | 1.0 (1) | 1.0 (1) | 1.1 (1) | 1.1 (1) | 1.1 (1) | 1.1 (1) |       |
| His  |       |      |      |      |         |      |        |     |
| Arg  | 1.1 (1) | 1.0 (1) | 0.7  | 1.0 (1) | 1.0 (1) | 1.0 (1) |       |
| Trp  | (1) |      |      |      |         |      |        |     |
| Asp  | 1.2 (1) | 1.1 (1) | 4.0 (4) | 1.0 (1) | 2.7 (2) | 2.0 (2) | 4.4 (4) | 2.1 (2) |
| Thr  | 4.0 (4) |      |      |      |         |      |        |     |
| Ser  | 1.0 (1) | 1.1 (1) | 4.7 (5) | 6.4 (5) | 5.0 (5) | 8.9 (10) |       |
| Gln  | 1.0 (1) | 1.0 (1) | 2.1 (2) | 0.9 (1) | 1.5 (2) | 2.2 (2) |       |
| Pro  |       |      |      |      |         |      |        |     |
| Gly  | 1.1 (1) | 2.3 (3) | 1.3 (1) | 4.1 (4) | 1.1 (1) |       |       |
| Ala  | 0.3  |      |      |      |         |      |        |     |
| Val  | 1.0 (1) | 1.9 (2) | 2.1 (2) | 0.3  | 2.7 (2) |       |       |
| Ile  | 1.0 (1) | 3.6 (3) | 2.6 (3) | 6.1 (6) |       |       |
| Leu  | 3.9 (4) | 1.8 (2) | 3.0 (2) | 2.6 (3) |       |       |
| Tyr  |       |      |      |      |         |      |        |     |
| Phe  | 1.9 (2) | 3.3 (5) | 1.0 (1) | 3.3 (5) | 1.0 (1) |       |       |
| Total | 6    | 3    | 20   | 14   | 40      | 16   | 56     | 9   |
| Yields | 25   | 73   | 40   | 36   | 8       | 5    | 47     | 50  |

a. Values reported are amino acid residues. Amino acids present at a level less than 0.2 residues are omitted. The assumed integral numbers of residues are given in parentheses.
b. Presence of tryptophan was determined by staining with p-dimethylaminobenzaldehyde.
c. Yields are based on micromoles of peptides isolated compared with micromoles of CNBr fragment F₃ digested with trypsin.

Fig. 2. Ion exchange chromatography of the soluble material from a tryptic digest of Fragment F₃ (10 µmol) on a column (1 × 30 cm) of DEAE-cellulose in 0.02 M Tris, pH 8.0. At the point indicated by the arrow (↓) a linear gradient of 0 to 0.3 M KCl was initiated. The solid line denotes the absorbance of the effluent at 230 nm. Each tube contained 3.0 ml. The letters A to J denote fractions pooled for purification; peptides (T11b to T17) obtained from each fraction are also noted.

Fig. 3. Gel filtration of the peptides insoluble in 0.02 M Tris, pH 8.0, from a tryptic digest of F₃ (10 µmol) on a column (2.2 × 100 cm) of Sephadex G-50 in 8 M urea, 1 M propionic acid. Each tube contained 3.0 ml of effluent. The absorbance of effluent fractions at 280 nm is designated by the solid line.
analyzed directly and cleavage into smaller peptides was re-
by high voltage paper electrophoresis (pH 6.5 and
sequence of tryptic peptides T11b, T12, and T17 (Table I) was
integral numbers of residues are given in parentheses.
peptide T13, T14, and T15, S8 to S10 extended the sequence of peptide
position 146 where no dansyl-amino acid was detected. The
minimal amount of time during each step of the degradation (18,
indicated that residue 146 was alanine and similar treatment of
material that was insoluble at pH 8.0 with succinic anhydride fol-
with nanomoles of T13 digested with subtilisin or after hydrolysis
of the tryptic peptide with 0.03 M HCl.

**Amino Acid Sequence of Tryptic Peptides**—The amino acid sequence of tryptic peptides T11b, T12, and T17 (Table I) was determined directly by Edman degradation and dapsylation (Fig. 1). Peptides T13, T14, and T15,16 were too large to be analyzed directly and cleavage into smaller peptides was required.

**Peptide T13**—Direct analysis of peptide T13 (Fig. 1) gave the sequence of the first 15 residues of T13, with the exception of position 146 where no dansyl-amino acid was detected. The residue at this position and the remainder of the sequence were established by digestion of peptide T13 with subtilisin as well as by partial acid hydrolysis of the tryptic peptide. The subtilisin peptides, S7 to S10 (Table II), were isolated by high voltage paper electrophoresis (pH 4.7 and 1.9). Edman analysis of S7 indicated that residue 146 was alanine and similar treatment of peptides S8 to S10 extended the sequence of peptide T13 to residue 156 (Fig. 1). The remaining sequence was established by the isolation and characterization of peptide PA1 (Table II). Peptide PA1 was one of a mixture of peptides obtained after partial acid hydrolysis (0.03 M HCl, 11 hours) and high voltage electrophoresis at pH 4.7. The amino acid sequence of peptide PA1 completed the determination of the amino acid sequence of the tryptic peptide T13 (Fig. 1).

**Peptide T14**—The major difficulty associated with the determination of the sequence of peptide T14 (Table I) stemmed from the fact that Edman degradation as routinely used released no new residues after removal of residue 161 (Fig. 1). This difficulty was circumvented by exposing the peptide to acid for a minimal amount of time during each step of the degradation (18, 25), which allowed extension of the sequence to residue 165. The sequence was confirmed by the isolation and analysis of peptides PA2 and PA3 (Table III). These peptides were obtained after treatment of T14 with 0.1 M HCl for 1 hour at 110° by high voltage paper electrophoresis (pH 6.5 and 1.9).

**Table II**

| Amino acids | S7 | S8 | S9 | S10 | PA1 |
|-------------|----|----|----|-----|-----|
| Arg         | 1.0 (1) | 2.1 (2) | 0.9 (1) | 0.8 (1) |
| Asp         | 0.8 (1) | 1.0 (1) | 1.0 (1) | 1.0 (1) |
| Thr         | 0.9 (1) | 0.8 (1) | 0.9 (1) | 1.2 (1) |
| Gly         | 1.0 (1) | 1.0 (1) | 1.0 (1) | 1.0 (2) |
| Ala         | 1.0 (1) | 1.0 (1) | 1.0 (1) | 1.0 (1) |
| Leu         | 1.0 (1) | 1.0 (1) | 1.0 (1) | 1.0 (1) |
| Total residues | 4 | 2 | 6 | 4 | 5 |
| Yield (%) | 7 | 12 | 24 | 7 | 20 |

* Values reported are amino acid residues. Amino acids present at a level less than 0.2 residue are omitted. The assumed integral numbers of residues are given in parentheses.

**Table III**

| Amino acids | PA1 | PA3 | S11 | Pm 4 |
|-------------|-----|-----|-----|------|
| Arg         | 1.1 (1) | 0.9 (1) | 0.9 (1) | 1.0 (1) |
| Asp         | 0.9 (1) | 1.0 (1) | 1.0 (1) | 1.0 (1) |
| Thr         | 1.1 (1) | 0.9 (1) | 1.8 (2) | 1.0 (1) |
| Gly         | 0.3 | 1.8 (2) | 1.1 (1) | 1.0 (1) |
| Val         | 1.0 (1) | 1.0 (1) | 1.0 (1) | 1.0 (1) |
| Total residues | 2 | 3 | 6 | 4 |
| Yield (%) | 5 | 4 | 24 | 24 |

* Values reported are amino acid residues. Amino acids present at a level less than 0.2 residue are omitted. The assumed integral numbers of residues are given in parentheses.

The amino acid sequence of the COOH-terminal portion of T14 (residues 166 to 172, Fig. 1) was determined by analysis of peptides S11 and Pm 4 (Table III). Peptide S11 was obtained from a subtilisin digest of T14 by high voltage electrophoresis at pH 4.7 followed by electrophoresis at pH 1.9. Peptide Pm 4 was isolated from a pronase digest of T14 by electrophoresis at pH 1.9. The amino acid sequences of these peptides completed the determination of the sequence of T14.

**Peptide T15,16**—The determination of the amino acid sequence of peptide T15,16 proved to be the most difficult task in the determination of the primary structure of Con A. As described above, the failure of trypsin to cleave at Lys 200 and the chymotryptic cleavage at Phe 212 resulted in a family of similar peptides of which only T16b could be obtained in homogeneous form. In addition, peptides T15,16a and T15,16 were insoluble at neutral pH and therefore were often resistant to digestion with other proteolytic enzymes.

Direct analysis of T15,16a or T15,16 (Table I) gave the sequence of residues 173 to 180 with the exception of positions 178 and 182, where no dansyl-amino acids were detected. These data indicated that the presence of heterogeneity in the isolated peptides would be associated with the COOH-terminal portion of the peptide. The remaining sequence of T15,16 was established by isolating and characterizing peptides obtained after digestion with other enzymes and after partial acid hydrolysis (Table IV) and by analyzing peptide T16b (Table I).

T15,16 was digested with pepsin and fractionated on DEAE-cellulose (Fig. 4). Material in Fractions C and E gave peptides P15 and P13, respectively. Peptide P14 was obtained from material in Fraction D by high voltage electrophoresis at pH 4.7 followed by electrophoresis at pH 1.9. Sequence analysis of P13 indicated that residue 178 is proline and confirmed the sequence of residues 176 to 185. The sequence of P14 provided the sequence of residues 192 to 195 and the partial sequence of P15 gave residues 199 to 210, with the exception of positions 200 and 209, where no dansyl-amino acids were detected.

Partial acid hydrolysis proved particularly valuable in the analysis of this region of the molecule. Peptides PA5, PA5a, PA6, and PA7 (Table IV) were obtained by hydrolysis (0.03 M HCl, 4 hours) of peptic peptide P15, followed by two-dimensional paper electrophoresis at pH 4.7 and 1.9. Analysis of these peptides (Fig. 1) established that residues 200 and 209 were
lysine and glycine, respectively, confirmed the sequence results on peptide P15 and extended the sequence of P15 to residue 212, the COOH terminus of T15,16a. Partial acid hydrolysis also proved valuable in filling the gap between residues 186 and 189. Peptide PA4 (Table IV) was isolated after partial acid hydrolysis (0.25 M acetic acid, 8 hours at 110°) of T15,16 by high voltage electrophoresis at pH 6.5 followed by paper chromatography.

The amino acid sequence of peptide PA4 (Fig. 1) extended the sequence from the COOH terminus of P13 to residue 189. The sequences of the remaining gaps in peptide T15,16a were provided by analysis of peptides obtained from digestion of the tryptic peptide with thermolysin and subtilisin.

The thermolysin peptides Th16 and Th17a (Table IV) were obtained from digests of T15,16 and peptides Th15 and Th17 (Table IV) were obtained from digests of succinylated T15,16. In both cases, gel filtration of these digests on Sephadex G-25 gave an elution profile similar to the one shown in Fig. 5. High voltage electrophoresis at pH 6.5 of material in Fraction D gave Th16 and of that in Fraction F gave Th17a. Peptides Th15 and Th17 were obtained from material corresponding to that in Fractions D and G, respectively. Peptide Th15a (Table IV) was obtained from a direct thermolysin digest of Fragment F3 by ion exchange chromatography on DEAE-cellulose. Peptide Th15a was subjected to hydrolysis with enzymes or KCI.

**Table IV**

| Amino Acid Composition of Peptides Used to Establish Sequence of T15,16a |
|---|
| **Residue** | **P13** | **P14** | **P15** | **PA4** | **PA5a** | **PA6** | **PA7** | **S12a** | **Th15** | **Th13a** | **Th13b** | **Th16** | **Th17** | **Th17a** | **Th18** |
| Lys | 1.1 | 1.0 | 0.6 | 0.9 | 1.0 | | | 1.1 | | | | | | | | | 0.8 |
| His | 1.0 | 1.0 | | | | | | 1.0 | | | | | | | | | |
| Arg | | | 1.0 | | | | | | | | | | | | | | | |
| Trp | | | | | | | | | | | | | | | | | | | 1.2 |
| Ser | 0.9 | 0.9 | 1.2 | 1.0 | 0.9 | | | 2.6 | 0.9 | 1.0 | 1.0 | | | | | | |
| Glu | 1.1 | 1.1 | 1.1 | 1.0 | 0.9 | | | 1.6 | 1.1 | 1.1 | 1.0 | | | | | | |
| Pro | 0.9 | 1.3 | 1.1 | 1.0 | 0.9 | | | 1.0 | | | | | | | | | |
| Gly | 0.3 | 1.1 | 2.0 | 1.1 | 1.0 | | | 0.7 | | | | | | | | | |
| Ala | 1.9 | 1.1 | 2.1 | 1.1 | 1.0 | | | 1.0 | | | | | | | | | |
| Val | 1.1 | 1.1 | | | | | | 1.0 | | | | | | | | | |
| Ile | 1.1 | 1.1 | | | | | | 0.3 | | | | | | | | | |
| Leu | 0.5 | 1.0 | 2.1 | 1.1 | | | | 1.0 | | | | | | | | | |
| Total residues | 11 | 14 | 14 | 14 | 14 | 14 | 14 | 14 | 14 | 14 | 14 | 14 | 14 | 14 | 14 | 14 |
| Yield (230 nm) | 13 | 23 | 6 | 1 | 6 | 17 | 7 | 16 | 15 | 3 | 11 | 14 | 20 | 5 | 1 | 40 |

a. Values presented are amino acid residues. Amino acids present at a level less than 0.2 residues are omitted. The assumed integral numbers of residues are given in parentheses.

b. The presence of tryptophan was determined by staining with p-dimethylaminobenzaldehyde.

c. The presence of tryptophan was assumed on the basis of composition.

d. The presence of tryptophan was determined by digestion with leucine aminopeptidase followed by amino acid analysis.

e. Yields are based on nanomoles of peptides isolated compared with nanomoles of CNBr fragment F3 or T15,16 subjected to hydrolysis with enzymes or KCI.

**Fig. 4.** Ion exchange chromatography of a peptic digest of T15,16 on a column (1 X 15 cm) of DEAE-cellulose in 0.02 M Tris, pH 8.0. The absorbance of effluent fractions at 230 nm is indicated by the solid line.

**Fig. 5.** Gel filtration of a thermolysin digest of T15,16 on a column (2.2 X 100 cm) of Sephadex G-25 in AmProP. Each tube contained 2.8 ml of effluent; the absorbance of effluent fractions at 230 nm is indicated by the solid line.
Treatment of peptide Th15a with leucine aminopeptidase released isoleucine, followed by a slower release of tryptophan, establishing residue 182 as tryptophan (Fig. 1). Treatment of Th15 with carboxypeptidase A indicated that residue 190 was alanine. Peptide Th15b was also treated with leucine aminopeptidase, which indicated that the NH2-terminal residue of this peptide was valine. Edman analysis of Th16, Th17, Th17a, and S12a completed the amino acid sequence of the portion of T15,16 represented by peptide T15,16a. The remaining sequence of T15,16 was established by analysis of peptide T16b (Table I).

Direct analysis of T16b yielded the sequence of the first 6 residues; the sequence of the peptide was completed by the isolation and characterization of peptide Th18. This peptide was obtained directly by gel filtration of a thermolysin digest of T16b on Sephadex G-25 (Fig. 6) and also by gel filtration of a thermolysin digest of T15,16 (Fig. 5, Fraction C) followed by high voltage electrophoresis at pH 6.5. Edman analysis of Th18 (Fig. 1) completed the determination of the amino acid sequence of T15,16 and the sequence of all the tryptic peptides of F3.

Order of Tryptic Peptides—Direct sequence analysis of CNBr Fragment F3 in the automatic sequenator placed peptides T11b, T12, and T13 (Fig. 1). This assignment was confirmed by Edman analysis of peptide S6a (Table V) which was obtained from a subtilisin digest of Fragment F3 by gel filtration on Sephadex G-25 in AmProl followed by ion exchange chromatography on DEAE-cellulose. The amino acid composition and partial sequence of peptide C8 (Table V) established the positions of peptides T14 and T15,16. Peptide C6 was obtained from a chymotryptic digest of Fragment F3 by gel filtration on Sephadex G-50 in AmProl followed by high voltage paper electrophoresis at pH 4.7, or alternatively, by ion exchange chromatography on DEAE-cellulose. The amino acid composition and partial sequence of this peptide indicated that it began at residue 157, spanned all of peptide T14, and extended into peptide T15,16. Treatment with carboxypeptidase A (Fig. 1) placed the COOH terminus of C6 at Phe 173, thereby confirming that it overlapped T14 and T15,16.

Peptide Th18a was obtained from a thermolysin digest of Fragment F3 by gel filtration first on Sephadex G-50, then on Sephadex G-25, and finally by ion exchange chromatography on DEAE-cellulose. The amino acid sequence (Fig. 1) of peptide Th18a (Table V) placed T17 adjacent to T15,16 at the COOH terminus of the polypeptide chain. This is consistent with the fact that the COOH terminus of T17 was identical with that of the Con A molecule (2, 17).

Asparaginyl and Glutaminyl Residues in F3—The assignment of the positions of all asparaginyl and glutaminyl residues was made on the basis of the electrophoretic mobilities and amino acids released after enzymatic digestion of peptides. Those peptides used in determining the amino acid sequence of F3 were described above and the amino acid compositions are given in Tables I to V. The amino acid compositions of the other peptides are listed in Table VI, and the sequence represented by each peptide is indicated in Fig. 1. The electrophoretic mobilities of all peptides are given in Table VII.

Peptide C5 (Tables VI and VII) was obtained from a chymotryptic digest of Fragment F3 by gel filtration on Sephadex G-50 in AmProl followed by high voltage paper electrophoresis at pH 6.5. This peptide was neutral at pH 6.5 and indicated that residue 131 was asparagine and residue 132 was glutamine. Peptide T12 (Table I) was also neutral at pH 6.5 (Table VI), and therefore, the assignment of Aex 136 and Glx 137 had to be made by other means. Digestion of T12 with aminopeptidase M gave 1.0 residue of aspartic acid, 1.0 residue of glutamic acid, and 1.0 residue of lysine, allowing the assignment of Asp 136 and Gln 137.

The assignment of the positions of asparaginyl and glutaminyl residues in peptide T13 was difficult because there were 6 such residues spaced closely within this peptide. The assignment of Asp 139 was made on the basis of the electrophoretic mobility (Table VII) of a peptide (C11, Table VI) obtained by electrophoresis at pH 6.5 from a chymotrypsin C digest of tryptic peptide T13. This finding also allowed the assignment of Gln 143 on the basis of the electrophoretic mobility of peptide S6b (Table VI). Peptide S6b and peptides S7,8 and S9,10 were isolated from subtilisin digests of peptide T13 by two-dimensional high voltage electrophoresis at pH 6.5 and 1.0. The electrophoretic mobility of peptide T13 was further confirmed by amino acid analysis (Table VII) of a peptide (C1, Table V) obtained by electrophoresis of a crude extract of C elegantus at pH 6.5 and 1.0. The electrophoretic mobility of peptide T13 was further confirmed by amino acid analysis (Table VII) of a peptide (C1, Table V) obtained by electrophoresis of a crude extract of C. elegantus at pH 6.5 and 1.0.

### Table V

| Amino acids | S6a | C6 | Th18a |
|-------------|-----|----|-------|
| Lys         | 1.7 (2) | 1.9 (2) | 1.0 (1) |
| Arg         | 2.1 (2) | 1.2 (1) | 1.3 (1) |
| Asp         | 0.8 (1) | 4.6 (5) | 3.6 (4) |
| Thr         | 2.2 (2) | 1.1 (1) |       |
| Ser         | 1.0 (1) | 0.7 (1) |       |
| Glu         | 3.0 (3) | 1.9 (2) |       |
| Pro         | 1.0 (1) | 2.1 (2) |       |
| Gly         | 2.1 (2) | 0.8 (1) | 0.9 (1) |
| Ala         | 0.5 (1) |       |       |
| Val         | 1.0 (1) |       |       |
| Ile         | 0.8 (1) |       |       |
| Leu         | 2.1 (2) |       |       |
| Total residues | 10 | 19 | 13 |
| Yield (%)  | 14 | 4 | 12 |

* Values presented are amino acid residues. Amino acids present at a level less than 0.2 residue are omitted. The assumed integral numbers of residues are given in parentheses.

* Yields are based on micromoles of peptides isolated compared with micromoles of CNBr Fragment F3 digested with subtilisin, chymotrypsin, or thermolysin.

![Fig. 6. Gel filtration of a thermolysin digest of T16b on a column (2.5 × 100 cm) of Sephadex G-25 in AmProl. Each tube contained 2.0 ml of eluent; the absorbance of eluent fractions at 230 nm is indicated by the solid line.](http://www.jbc.org/)
TABLE VI

Amino acid compositions of peptides used to establish positions of asparaginyl and glutaminyl residues in FP

| Amino acids | C5 | C7 | C8 | Pre 6 | Pre 7 | Pre 8 | Pre 8a | P10 | Asp | Thr | Ser | Glu | Pro | Gly | Ala | Ile | Leu | Phe |
|-------------|----|----|----|-------|-------|-------|-------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Lys         | 0.9 (1) |     |     |       |       |       |       |     |     |     |     |     |     |     |     |     |     |     |
| His         | 1.0 (1) | 0.9 (1) |     |       |       |       |       |     |     |     |     |     |     |     |     |     |     |     |
| Arg         | 0.8 (1) |     |     |       |       |       |       |     |     |     |     |     |     |     |     |     |     |     |
| Thr         | 1.0 (1) |     |     |       |       |       |       |     |     |     |     |     |     |     |     |     |     |     |
| Ser         | 0.3 |     |     |       |       |       |       |     |     |     |     |     |     |     |     |     |     |     |
| Glu         | 0.8 (1) |     |     |       |       |       |       |     |     |     |     |     |     |     |     |     |     |     |
| Pro         |     |     |     |       |       |       |       |     |     |     |     |     |     |     |     |     |     |     |
| Gly         | 0.3 |     |     |       |       |       |       |     |     |     |     |     |     |     |     |     |     |     |
| Ala         | 1.1 (1) |     |     |       |       |       |       |     |     |     |     |     |     |     |     |     |     |     |
| Ile         |     |     |     |       |       |       |       |     |     |     |     |     |     |     |     |     |     |     |
| Leu         | 1.0 (1) |     |     |       |       |       |       |     |     |     |     |     |     |     |     |     |     |     |
| Phe         | 2.0 (2) |     |     |       |       |       |       |     |     |     |     |     |     |     |     |     |     |     |
| Total residu- | 4 | 4 | 7 | 4 | 6 | 7 | 5 | 3 | 2 | 4 | 5 | 6 | 9 | 10 | 4 | 6 | 5 | 3 | 3 | 4 |
| Yield (%)    | 30 | 4 | 24 | 6 | 11 | 4 | 23 | 3 | 43 | 33 | 19 | 38 | 24 | 18 | 10 | 11 | 11 | 19 |

* Values presented are amino acid residues. Amino acids present at a level less than 0.2 residue are omitted. The assumed integral numbers of residues are given in parentheses.

Yields are based on nanomoles of peptides isolated compared with nanomoles of tryptic peptides or CNBr Fragment Fr digested with various enzymes.

TABLE VII

Electrophoretic mobilities of peptides used to determine positions of asparaginyl and glutaminyl residues in FP

| Residue | Peptide | Ref Table | Mol Wt | Mobility | Net Charge |
|---------|---------|-----------|--------|----------|------------|
| Asn 131 | C5 (Phe-Asx-Glx-Phe) | VI | 556.6 | 0 | 0 |
| Gln 132 | | | | | |
| Asp 136 | T12 (Asx-Glx-Lys) | I | 390.4 | 0 | 0 |
| Gln 177 | | | | | |
| Asp 139 | CCl (Asx-Leu-Ile) | VI | 359.5 | 0.49 | -1 |
| Gln 143 | S6 (Asx-Leu-Ile-Leu-Glx) | VI | 472.6 | 0.46 | -1 |
| Asp 145 | S7,8 (Gly-Asx-Ala-Thr-Thr-Gly) | VI | 520.3 | 0.44 | -1 |
| Asp 151 | S9 (Thr-Asx-Gly-Leu-Glx) | II | 648.6 | 0.91 | -2 |
| Gln 153 | S9,10 (Thr-Asx-Gly-Leu-Glx-Leu-Thr-Arg) | VI | 1019.7 | 0.27 | -1 |
| Asp 155 | SE1 (Asx-Leu-Glx) | VI | 375.4 | 0.50 | -1 |
| Gln 155 | SE2 (Gly-Ser-Pro-Glx) | VI | 388.3 | 0.36 | -1 |
| Gln 183 | TH15a (Ile-Thr-Glx-Ser) | IV | 533.6 | 0.32 | -1 |
| Glu 192 | C7 (Phe-Glx-Ala-Thr) | VI | 466.6 | 0.45 | -1 |
| Asp 203 | P15 (Lys-Ser-Pro-Asx) | VI | 445.5 | 0 | 0 |
| Gln 206 | | | | | |
| Asp 216 | S Prn 1 (Ser-Asx-Gly-Ser-Pro-Glx) | VI | 590.5 | 0.39 | -1 |
| Asp 218 | S Prn 2 (Asx-Gly-Ser-Pro-Glx) | VI | 530.3 | 0.44 | -1 |
| Asp 235 | PRN 3 (Asx-Gly-Ile) | VI | 503.3 | 0.63 | -1 |
| Asp 237 | Pn 8a (Asx-Gly) | VI | 190.2 | 0.78 | -1 |
| Asp 238 | S12 (Ser-Pro-Asx-Ser-His-Pro-Asx-Gly-Ile) | VI | 995.0 | 0.34 | -1 |
| Asp 239 | PRN 7 (Ser-His-Pro-Asx-Aux-Gly) | VI | 582.6 | 0.39 | -1 |
| Asp 246 | PRN 8a (Asx-Gly) | VI | 190.2 | 0.78 | -1 |
| Asp 245 | PRN 8b (Asx-Gly) | VI | 190.2 | 0.78 | -1 |
| Asp 247 | PRN 8c (Asx-Gly) | VI | 190.2 | 0.78 | -1 |

* Mobilities relative to aspartic acid after electrophoresis at pH 6.5; net charge calculated by the method of Offord (1966).

The mobility of peptide S7,8 (Tables VI and VII) allowed the assignment of Asp 145. The assignments of Asp 139, Gln 143, and Asp 145 were confirmed by digestion of peptides S6b and S7,8 with aminopeptidase M followed by amino acid analysis (Table VI). Peptide S6b gave 1 residue each of aspartic acid and glutamine and S7,8 yielded 1 residue of aspartic acid.

The remaining three positions (Asx 151, Asx 153, and Glx 155) in T13 were contained within peptide S9,10 (Table VI), which included an arginine and had a net charge of -1, and peptide S9 (Table II) which contained no basic residues and had a net charge of -2 at pH 6.5. These data indicate that at least two of these positions were occupied by acidic residues. Peptide SE1 was isolated by paper electrophoresis at pH 6.5 of peptide S9 after three steps of the Edman degradation. This peptide
lacked residue 151 and had a net charge of -1. This result showed that either Asx 153 or Glx 155, but not both, was an acidic residue, and that residue 151 was aspartic acid. Peptide P19 (Table VI), obtained from a peptic digest of peptide T13 by ion exchange chromatography on DEAE-cellulose, contained arginine and was neutral (Table VII) at pH 6.5, indicating that residue 155 was glutamic acid. Residue 153 therefore must be asparagine.

Within peptide T14 the electrophoretic mobilities (Table VII) of peptides S Pm 1, S Pm 2, and SE2 (Table VI) allowed the assignments of Asn 162 and Glu 166. Peptides S Pm 1 and S Pm 2 were obtained after digestion of tryptic peptide T14 first with subtilisin and then with pronase, followed by two-dimensional paper electrophoresis at pH 4.7 and 1.9. Peptide SE2 was obtained by two-dimensional paper electrophoresis (pH 6.5 and 1.9) of peptide S Pm 1 after two steps of the Edman degradation.

Within the region represented by tryptic peptide T15,16, Glu 183 and Glu 192 were established on the basis of the electrophoretic mobilities of peptide Th15a (Table IV) and peptide C7 (Table VI). Peptide C7 was obtained from a chymotryptic digest of Fragment F3 by gel filtration on Sephadex G-50 in AmProP followed by electrophoresis at pH 6.5. Similarly Asp 203 was assigned on the basis of the electrophoretic mobilities of peptides Pm 5 and Pm 6 (Table VI); these data and the electrophoretic mobilities of peptides Pm 7, Prri 8, Pm 8a, and S12 indicated that residue 208 was aspartic acid. Peptides Pm 5 through Pm 8a were obtained from pronase digests of peptide T15,16 by two-dimensional paper electrophoresis (pH 6.5 and 1.9), and peptide S12 was isolated from a subtilisin digest of peptide T15,16 by gel filtration on Sephadex G-25 in AmProP followed by electrophoresis at pH 6.5.

The assignment of Asn 216 was made on the basis of the electrophoretic mobility of peptide S13 (Table VI), which was obtained from a subtilisin digest of peptide T16b by two-dimensional paper electrophoresis (pH 6.5 and 1.9). The electrophoretic mobility of peptide Th18 (Table IV) suggested that residue 218 was aspartic acid. This assignment was confirmed by digestion of the peptide with aminopeptidase M, which released 1 residue of isoleucine, 1 residue of aspartic acid, and 2 residues of serine.

There are 2 aspartyl or asparaginyl residues in peptide T17. The electrophoretic mobility (Table VII) of peptide C8 (Table VI), obtained in the same manner as described above for peptides C6 and C7, suggested that either residue 235 or 237 was aspartic acid. The COOH-terminal residue has been reported previously as asparagine (2); therefore, residue 235 was assigned as aspartic acid.

**DISCUSSION**

The complete amino acid sequence of Con A is presented in Fig. 7. This sequence differs from the tentative sequence we reported previously (28) in the deletion of an aspartyl or asparaginyl residue at position 72, in the reassignment of residue 87 as glutamic acid, and in two transpositions, one between Ile 106 and Leu 107 and the other between Trp 109 and Phe 111 (1). In addition, three regions of the polypeptide chain (residues 163 to 166, 187 to 190, and 196 to 198), where the sequences had been identified tentatively by interpreting the electron density map,
have now been established chemically. Any remaining ambiguities in the sequence data are contained in the regions represented by peptide T15, 16, particularly residues 157 to 190 and 196 to 198. However, the residues assigned by the chemical sequence are in excellent agreement with the trace of the polypeptide chain in this region of the electron density map. Finally, this study completes the assignment of the positions of all asparaginyl and glutaminyl residues.

Previous structural studies have shown that the molecular weight of the Con A polypeptide chain is about 26,000 (2, 3) and that Con A contains no covalently bonded carbohydrate, lipid, or other detectable prosthetic group (29, 30). The amino acid sequence of the 237 residues shown in Fig. 7, therefore, accounts for the entire Con A protomer, in agreement with the contents of the crystallographic asymmetrical unit (31, 32).

Examination of the sequence of Con A shows that it is in general agreement with the amino acid compositions reported for the individual CNBr fragments (17) as well as for the intact polypeptide chain (2, 3). In addition, there is no evidence for the presence of unusual amino acids such as canavanine, known to be present in jack beans (23). While a number of difficulties were encountered in the sequence determination, the sequence reported here is in excellent agreement with the x-ray crystallographic data (15, 16).

More detailed analysis of the sequence shows a striking distribution of certain amino acids in Con A (Fig. 8). The distribution of charged residues is generally more dense in the NH2-terminal half of the polypeptide chain than in the COOH-terminal portion (Fig. 8A). There are three clusters of charged groups which stand out: a group of 6 negatively charged residues between Asp 2 and Asp 28, a closely spaced group of positive charges between Lys 30 and Arg 60, and another cluster of negative charges between Asp 71 and Glu 87. The majority of these charged clusters are exposed on the surface of the Con A protomer (10) with the interesting exception of the first negatively charged cluster, in which 3 residues (Glu 8, Asp 10, and Asp 19) are involved in metal ion binding (Fig. 8A) (1, 15, 28). In contrast to the NH2-terminal portion, there are fewer charged residues in the COOH-terminal portion, and the charged residues are usually counterbalanced by the presence of oppositely charged residues in their vicinity. These observations seem to delineate the linear sequence of the Con A molecule into two distinct regions, arbitrarily separated at about residue 110.

This division between the two halves of the Con A protomer is even more striking when the distribution of aromatic amino acids is studied (Fig. 8B). Con A contains 4 tryptophans, 7 tyrosines, and 11 phenylalanines. The tryptophanyl residues are evenly distributed. In contrast, 6 of the 7 tyrosines are located in the NH2-terminal half and all of the phenylalanines are between residue 111 and the COOH terminus of the molecule (Fig. 8B). The predominance of such hydrophobic residues as well as the sparse distribution of charged residues in the last half of the linear sequence may account for the strongly aggregating behavior of CNBr Fragment F3 and of tryptic peptide T15, 16. In accord with these observations, examination of the x-ray crystallographic model of Con A indicates that these hydrophobic residues form a cluster in the interior of the molecule (16, 28). In addition, within the sequence of Fragment F3, residues 130 to 132, 136 to 139, and 175 to 178 all appear to be involved in noncovalent interactions between halves of the ellipsoidal dimers and, therefore, are relatively shielded from the solvent (16).

These features of the linear structure of Con A may prove particularly valuable for comparison with other mitogenic and nonmitogenic lectins, although it is clear that such features may not be readily apparent from partial sequence data. Furthermore, analysis of the relationship between primary and three-dimensional structure is obviously much more complex than is implied by the simple distribution of certain amino acid residues within the linear structure. Detailed features of the folding of the Con A polypeptide chain are described by Reeke et al. (16). Nevertheless, the fact that both the complete amino acid sequence and the three-dimensional structure of Con A have now been reported in detail (1, 15, 16) makes this protein a valuable source of new data for studying the factors that might be involved in predicting tertiary structure from the amino acid sequence of a protein. The tentative sequence of Con A (28) has already stimulated an analysis of the influence of nearest neighbor residues on the conformation of the intermediate amino acid residues (34), the development of empirical rules for predicting the initiation and termination of helical and β structures (35), and the analysis of nucleation points for protein folding (36).

Above all, however, the sequence should be particularly useful in correlating the modification of specific amino acid residues of chemically derivatized Con A with corresponding alterations in biological activity. As we have discussed in the previous paper (1), Lys 114 and Lys 116 have been implicated as residues affected by succinylation and acetylation of Con A (9), resulting in dissociation of the tetramer into dimers and alteration of the effects of the protein on lymphocyte surface receptors. Similar mechanisms may apply to the dissociation of the tetramer by maleylation (10).

In contrast to these chemical derivatives in which carbohydrate-binding specificity is preserved, reaction of Con A with N-acetylimidazole results in loss of carbohydrate-binding capacity, thus implicating the participation of tyrosyl residues in the binding of saccharides (11, 37). It has also been shown that derivatization of carboxyl groups using glycine methyl ester and carbodiimide resulted in inactivation of the carbohydrate-binding activity of Con A (8). This and the results of potentiometric titration studies have suggested that a carboxyl group is involved in the saccharide binding site of the lectin. In view of the facts that saccharide binding is dependent upon metal binding (38) and that the protein ligands in the metal binding site are predominantly carboxyl groups (1, 15, 28), the loss in carbohydrate-binding activity on derivatization with glycine methyl ester could occur as a result of the abolition of the metal-binding capacity of the protein rather than through direct action at the carbohydrate binding site. In any case, similar studies using specific affinity-labeling techniques (39) will now be particularly valuable in the precise location and detailed description of the saccharide binding site.
Finally, knowledge of the structure of Con A and the chemical characterization of its cell surface receptors should help in the design of experiments to localize other key features of the protein responsible for mitogenesis and cell membrane alteration in lymphoid and neoplastic cells.

Acknowledgments—We thank Doctors Joseph Becker and George Reeke for helpful discussions during the course of this work. We also thank Miss Katherine Bergmann, Mrs. Helvi Hjelt, Miss Catherine Volin, Mrs. Martha O'Connor, and Mrs. Joan Sadow for excellent technical assistance, and Miss Lucille Fallon for her help in preparing the manuscript.

REFERENCES
1. WANG, J. L., CUNNINGHAM, B. A., WAXDAL, M. J., AND EDELMAN, G. M. (1975) J. Biol. Chem. 250, 1513-1524
2. WAXDAL, M. J., WANG, J. L., PFLUMM, M. N., AND EDELMAN, G. M. (1971) Biochemistry 10, 3343-3347
3. CUNNINGHAM, B. A., GOTTLEIB, P. D., KONIGSBERG, W. H., AND EDELMAN, G. M. (1968) Biochemistry 7, 1983-1985
4. CUNNINGHAM, B. A., RUTTSHAUSER, U., GALL, W. E., GOTTLEIB, P. D., WAXDAL, M. J., AND EDELMAN, G. M. (1970) Biochemistry 9, 3161-3170
5. CUNNINGHAM, B. A., WANG, J. L., BERGEVARD, I., AND PETERSON, P. A. (1973) Biochemistry 12, 4811-4822
6. SCHWARTZ, J. H., AND EDELMAN, G. M. (1965) J. Exp. Med. 118, 41-51
7. ORFORD, R. E. (1966) Nature 211, 591-593
8. MILSTEIN, C., MILSTEIN, C. P., AND FEINSTEIN, A. (1969) Nature 221, 151-154
9. EDELMAN, G. M., CUNNINGHAM, B. A., REEKE, G. N., JR., BECKER, J. W., WAXDAL, M. J., AND WANG, J. L. (1972) Proc. Nat. Acad. Sci. U. S. A. 69, 2580-2584
10. OSLER, M. O. J., AND LIENER, I. E. (1967) Biochemistry 6, 105-111
11. AGRAWAL, B. B., AND GOLDSTEIN, I. J. (1968) Biochem. Biophys. Acta 145, 205-209
12. HASSING, G. S., GOLDSTEIN, I. J., AND MARINI, M. (1971) Biochim. Biophys. Acta 243, 90-97
13. GUNTHER, G. R., WANG, J. L., YAHARA, I., CUNNINGHAM, B. A., AND EDELMAN, G. M. (1973) Proc. Nat. Acad. Sci. U. S. A. 70, 1473-1477
14. CHOU, P. Y., AND FASMAN, G. D. (1974) Biochemistry 13, 222-225
15. NAGANO, K. (1974) J. Mol. Biol. 84, 337-372
16. DOYLE, R. J., AND ROHOLT, O. A. (1968) Life Sci. 7, 841-846
17. KALB, A. J., AND LEVITZKI, A. (1968) Biochem. J. 109, 669-672
18. THOMAS, E. W. (1973) Carbohydr. Res. 31, 101-107
The covalent and three-dimensional structure of concanavalin A. II. Amino acid sequence of cyanogen bromide fragment F3.
B A Cunningham, J L Wang, M J Waxdal and G M Edelman

J. Biol. Chem. 1975, 250:1503-1512.

Access the most updated version of this article at http://www.jbc.org/content/250/4/1503

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 0 references, 0 of which can be accessed free at http://www.jbc.org/content/250/4/1503.full.html#ref-list-1