

## The Primary Structure of Soybean Leghemoglobin

### II. Amino Acid Sequence of Seventeen Peptides Isolated from the Tryptic Hydrolysate of the Slow Component

NILS ELLFOLK and GUNNEL SIEVERS

*Department of Biochemistry, University of Helsinki, SF-00170 Helsinki 17, Finland*

The amino acid sequences of seventeen peptides which had been isolated from the tryptic hydrolysate of the apoprotein of the slow component (Lba) of soybean leghemoglobin have been determined. The amino terminal peptide has been identified as *Val-Ala-Phe-Thr-Glu-Lys*.

In the preceding paper<sup>1</sup> the purifications and amino acid composition of seventeen peptides from the tryptic digest of the slow components of soybean leghemoglobin (Lba) have been described. All these peptides were obtained in a homogeneous state and in sufficient quantity to allow the structural studies to be performed. The determination of the sequence of these peptides is described in this communication.

#### MATERIALS AND METHODS

*Leghemoglobin.* The slow component of soybean leghemoglobin (Lba) and its apoprotein were prepared as previously described.<sup>2,3</sup>

*Peptides.* As described in the preceding paper,<sup>1</sup> each peptide isolated was finally dissolved in 2 ml of deionized water and stored at  $-16^{\circ}\text{C}$ . Aliquots were removed and taken to dryness just before use for the structural studies.

*Enzymes.* Carboxypeptidase A was a crystallized, DFP-treated preparation and subtilo-peptidase A a crystallized preparation, both from Sigma Chemical Company (St. Louis, U.S.A.). Three times crystallized chymotrypsin and leucine aminopeptidase were preparations from Worthington Biochemical Corporation (Freehold, New Jersey, U.S.A.). Thermolysin was a crystalline preparation from Merck AG (Germany).

*Digestion of the peptides with enzymes.* Hydrolysis with chymotrypsin was carried out on 0.2 to 0.5  $\mu\text{mol}$  of peptides in 0.2–0.5 ml of 1% ammonium bicarbonate solution of pH 8.1 at  $37^{\circ}\text{C}$  for the appropriate time by addition of chymotrypsin at an enzyme-substrate ratio 1:100 (w/w). After digestion the solution was taken to dryness *in vacuo*.

*Abbreviations.* 1-Dimethyl-amino-5-naphthalenesulfonyl-, dansyl-, DNS-; 3-phenyl-2-thiohydantoin-, PTH-.

The digest with subtiloypeptidase A was performed on 0.2–0.5  $\mu\text{mol}$  of peptide in 0.2 M *N*-ethylmorpholine-acetic acid buffer, pH 8.0, containing 0.5  $\mu\text{mol}$  peptide/ml with a 1:100 molar ratio, at 37°C for the appropriate time. The reaction was terminated by holding the hydrolysate at 100°C for 3 min after which it was lyophilized.

Hydrolysis with thermolysin was performed in a 0.2 % ammonium carbonate solution, pH 8.3, containing 1  $\mu\text{mol}$  of peptide in 1 ml of solution at 37°C with a 1:100 molar ratio for 20 h, after which the digest was lyophilized.

*Partial hydrolysis with dilute hydrochloric acid.* 0.2–0.5  $\mu\text{mol}$  of peptide was hydrolyzed in 0.5 ml of 0.03 N HCl in sealed tubes for 15 h at 105°C. After hydrolysis the solution was taken to dryness *in vacuo* over pellets of sodium hydroxide.

*Partial hydrolysis with concentrated hydrochloric acid.* A sample of peptide solution containing 0.1  $\mu\text{mol}$  of peptide was evaporated to dryness, dissolved in 0.2 ml of 12 N hydrochloric acid and incubated at 37°C for 15–42 h. After hydrolysis the solution was taken to dryness *in vacuo* over pellets of sodium hydroxide.

*Chromatography of peptides.* The column chromatography and paper chromatography employed for the isolation and purification of the peptides were carried out as described in the preceding paper.<sup>1</sup> The chromatographic mobility was determined relative to leucine and denoted by  $R_{Lcu}$ .

*High voltage paper electrophoresis.* Purification of peptides by electrophoresis was carried out on Whatman No. 3MM paper in cooled tanks<sup>4</sup> at 60 V/cm. The electrophoresis buffers were pyridine-acetic acid at pH 6.5 and 3.5, 8 % acetic acid–2 % formic acid at pH 1.9, and ammonium carbonate (1 %) at pH 8.9.<sup>5</sup> The electrophoretic mobilities at pH 6.5 were determined relative to lysine and aspartic acid and are denoted by  $E_{Lys}$  and  $E_{Asp}$ , respectively.

*Amino acid analyses.* The amino acid content of the peptides was determined after hydrolysis at 108°C for 18 h on an automatic amino acid analyzer (Beckman/Spinco 120 B).<sup>6</sup> The general method of Spackman *et al.*<sup>6</sup> was used for determining the glutamine or asparagine content of peptides hydrolyzed with leucine aminopeptidase or carboxypeptidase A. The elution program used did not separate glutamine, asparagine, and serine. Tryptophan was quantitatively determined according to Spies and Chambers' method K.<sup>7</sup>

*Amino acid sequence analyses.* The phenylisothiocyanate method of Sjöquist *et al.*<sup>8</sup> and Blombäck *et al.*<sup>9</sup> was used. The thiohydantoin obtained were identified by thin layer chromatography in the solvent systems of Jeppsson and Sjöquist.<sup>10</sup> Subtractive Edman degradation was performed according to Konigsberg and Hill.<sup>11,12</sup> Edman-dansyl-degradation was performed according to Gray and Hartley,<sup>13,14</sup> and the DNS-amino acids were identified by electrophoresis at pH 4.38 at 4800 V for 2.5 h on a high voltage cold plate (Shandon Southern, Great Britain).<sup>13,14</sup> The DNS derivatives of serine, alanine, and glycine were further separated at pH 2.0 as described by Milstein.<sup>15</sup> The DNS derivatives were also identified by thin layer chromatography according to Woods and Wang.<sup>16</sup>

*Amide residues.* Amide residues were assigned on the basis of electrophoretic mobilities of the peptides at pH 6.5 according to Offord<sup>17</sup> and by digestion with leucine aminopeptidase.

*Hydrazinolysis.* For hydrazinolysis the peptide sample was dried thoroughly over  $P_2O_5$  and traces of moisture were removed by addition of 0.15 ml of benzene and repeating the drying procedure. The dry sample was sealed in an evacuated tube together with 0.1 ml of anhydrous hydrazine and heated for 16 h at 80°C. The sample was taken to dryness *in vacuo* over  $P_2O_5$  and  $H_2SO_4$ , and the liberated amino acid was determined with the amino acid analyzer.

*Digestion with carboxypeptidase A.* The incubation of 0.05–0.1  $\mu\text{mol}$  of peptide was performed in 0.2 ml of 0.2 M sodium bicarbonate buffer, pH 8.3, as previously described.<sup>18</sup> The liberated amino acids were determined with the automatic amino acid analyzer.

*Digestion with leucine aminopeptidase.* The incubation of 0.05–0.1  $\mu\text{mol}$  of peptide was performed as previously described in 0.1 ml of 0.1 M Tris-HCl buffer of pH 8.5.<sup>18</sup> Liberated amino acids were determined using the automatic amino acid analyzer.

*Nomenclature of leghemoglobin.* Leghemoglobin was abbreviated Lb and the different components of leghemoglobin as previously described.<sup>2</sup>

*Nomenclature of peptides.* The principles employed for numbering the peptides are identical with those used previously.<sup>1</sup> The letter before the peptide number indicates

the hydrolysis procedure used: "T" refers to trypsin, "Th" to thermolysin, "C" to chymotrypsin, "S" to subtilopeptidase A, "A" to hydrolysis with dilute and "H" with 12 N hydrochloric acid.

## RESULTS

*The amino-terminal sequence of Lba.* Amino-terminal valine was estimated in high yield as its PTH-derivative, as was the second amino acid alanine. The third, phenylalanine, was also conclusively identified, and the fourth amino acid seemed to be threonine. However, the fifth amino acid residue from the amino-terminus could not be clearly identified. The amino terminal of Lba is therefore *Val-Ala-Phe-Thr*.

*The sequence of the individual peptides.* The sequences shown in Table 1 were obtained from enzymatic digestions and Edman degradations of the

Table 1. Amino acid sequence of the tryptic peptides from the slow component (Lba) of soybean leghemoglobin.

| Peptide number    | Amino acid sequence   |
|-------------------|---|
| aT1 <sup>a</sup>  | Lys   |
| aT2               | Ala-Pro-Ala-Ala-Lys   |
| aT3               | Leu-Phe-Ala-Leu-Val-Arg   |
| aT4               | Ala-Lys   |
| aT5               | Thr-Ile-Lys   |
| aT6               | Ala-Ser-Gly-Thr-Val-Val-Ala-Asp-Ala-Ala-Leu-Gly-Ser-Val-His-Ala-Gln-Lys |
| aT7               | Ala-Val-Thr-Asn-Pro-Glu-Phe-Val-Val-Lys                                 |
| aT8               | Ala-Ala-Val-Gly-Asp-Lys   |
| aT9               | Asp-Ser-Ala-Gly-Glu-Leu-Lys   |
| aT10              | Glu-Ala-Leu-Leu-Lys   |
| aT11              | Leu-Thr-Gly-His-Ala-Glu-Lys   |
| aT12              | Val-Ala-Phe-Thr-Glu-Lys   |
| aT13 <sup>b</sup> | Glp-Asp-Ala-Leu-Val-Ser-Ser-Ser-Phe-Glu-Ala-Phe-Lys                     |
| aT14              | Asp-Leu-Phe-Ser-Phe-Leu-Ala-Asn-Pro-Thr-Asp-Gly-Val-Asn-Pro-Lys         |
| aT15              | [Glu-Asn-Ala-Leu] - [Val-Ser-Ser] - [Ser, Tyr] - [Glu-Ala-Tyr] - Lys    |
| aT16              | Trp-Ser-Asp-Glu-Leu-Ser-Arg   |
| aT17              | Ala-Trp-Glu-Val-Ala-Tyr-Asp-Glu-Leu-Ala-Ala-Ala-Ile-Lys                 |
| aT18              | Ala-Asn-Ile-Pro-Gln-Tyr-Ser-Val-Val-Phe-Tyr-Thr-Ser-Ile-Leu-Glu-Lys     |

<sup>a</sup> Assumed to be formed by secondary splitting of the peptides or free lysine contaminating the apoprotein. <sup>b</sup> Glp, abbreviation used for pyrrolidone-carboxylic acid.

tryptic peptides of component *a* of soybean leghemoglobin (Lba). The sequence of the histidine containing heptapeptide, aT11, published earlier,<sup>18</sup> is also included in the table. The peptides are discussed in the order in which they were originally isolated. The exact amino acid composition has been reported previously.<sup>1</sup> The amino acid composition of the peptides before and after each stage of the Edman degradation is expressed as mol ratios. Im-

purities generally amounted to 0.1 or less residues per mol and were not reported. Peptides which contain a lysine or arginine residue are assumed to be COOH-terminal.

*Peptide aT1.* This was identified as free *lysine* by paper electrophoresis, paper chromatography and by its position on the amino acid analyzer.

*Peptide aT2.* The amino acid sequence of this peptide was deduced from two steps of the Edman degradation presented in Table 2.

*Peptide aT3.* This peptide was degraded by the Edman technique through four steps, giving the sequence shown in Table 2.

Table 2. Amino acid sequences of peptides *aT2*, *aT3*, and *aT5*.

|                     |                                   |      |      |      |      |              |
|---------------------|-----------------------------------|------|------|------|------|--------------|
| Sequence <i>aT2</i> | Ala - Pro - Ala - Ala - Lys       |      |      |      |      |              |
| Edman degradation   |                                   |      |      |      |      |              |
| Step 1              | 0.00                              | 1.00 | 2.14 |      |      | <sup>a</sup> |
| Step 2              | 0.00                              | 0.00 | 2.0  |      |      | <sup>a</sup> |
| Sequence <i>aT3</i> | Leu - Phe - Ala - Leu - Val - Arg |      |      |      |      |              |
| Edman degradation   |                                   |      |      |      |      |              |
| Step 1              | 0.00                              | 1.00 | 1.06 | 0.97 | 0.82 | <sup>a</sup> |
| Step 2              | 0.00                              | 0.07 | 0.78 | 1.23 | 0.93 | <sup>a</sup> |
| Step 3              | —                                 | 0.00 | 0.34 | 1.03 | 0.93 | <sup>a</sup> |
| Step 4              | —                                 | —    | 0.00 | 0.39 | 1.00 | <sup>a</sup> |
| Sequence <i>aT5</i> | Thr - Ile - Lys                   |      |      |      |      |              |
| Edman degradation   |                                   |      |      |      |      |              |
| Step 1              | 0.00                              | 1.00 |      |      |      | <sup>a</sup> |

<sup>a</sup> Not determined.

*Peptide aT4.* From the specificity of trypsin, the dipeptide had to be *Ala-Lys*. In confirmation, the amino terminal endgroup was found to be alanine identified as PTH-alanine.

*Peptide aT5.* One step of Edman degradation (Table 2) showed threonine to be the NH<sub>2</sub>-terminal amino acid of this tripeptide.

*Peptide aT6.* The sequence of this peptide has earlier been reported.<sup>18</sup> The amino acids of positions 5–7 of this peptide were re-investigated. A thermolysin digestion of the peptide yielded four peptides, Th1, Th2, Th3, and Th4, which were isolated by electrophoresis at pH 6.5 and 1.9. Peptide Th2 was a pentapeptide containing two valine, one alanine, and one aspartic acid. It was subjected to three steps of Edman degradation, which formulated the sequence as *Val-Val-Ala-Asp*, (Table 3), a sequence identical to that reported by Aggarwal and Riggs.<sup>19</sup>

*Peptide aT7.* Five steps of Edman degradation and digestion with leucine aminopeptidase yielded the NH<sub>2</sub>-terminal sequence *Ala-Val-Thr-Asn-Pro* for this peptide (Table 4). When *aT7* was subjected to hydrolysis with subtilopeptidase A, four principal peptides were formed, S1, S2, S3, and S4.

Table 3. Amino acid sequence of peptide  $\alpha T6$ .

| Sequence                | Ala-Ser-Gly-Thr-Val-Val-Ala-Asp-Ala-Ala-Leu-Gly-Ser-Val-His-Ala-Gln-Lys |
|-------------------------|---|
|                         | ----- Th1 ----- ----- Th2 ----- ----- Th3 ----- ----- Th4 -----         |
| Thermolytic peptides    |   |
| Th1 (neutral)           | Ala, 1.05; Ser, 1.01; Gly, 1.07; Thr, 0.87                              |
| Th2 ( $E_{Asp}$ , 0.39) | Val, 1.38; Ala, 3.20; Asp, 1.42   |
| Edman degradation       |   |
| Step 1                  | Val, 0.93; Ala, 2.93; Asp, 1.13   |
| Step 2                  | Val, 0.00; Ala, 2.85; Asp, 1.15   |
| Step 3                  | — Ala, 1.90; Asp, 1.10  |
| Step 4                  | — Ala, 2.00; Asp, 0.68  |
| Th3 (neutral)           | Leu, 0.90; Gly, 1.09; Ser, 1.01   |
| Th4 ( $E_{Lys}$ , 0.74) | Val, 0.75; His, 0.40; Ala, 1.13; Glu, 1.04; Lys, 1.07                   |

Table 4. Amino acid sequence of peptide  $\alpha T7$ .

| Sequence                             | Ala-Val-Thr-Asn-Pro-Glu-Phe-Val-Val-Lys  |
|--------------------------------------|--|
|                                      | ----- S1 -----  S2  ----- S3 -----   |
|                                      | ----- Th1 ----- ----- Th2 ----- ----- S4 -----   |
|                                      | ----- Th3 -----  |
| Edman degradation                    |  |
| Step 1                               | Ala, 0.00; Val, 2.73; Thr, 0.99; Asp, 1.08; Pro, 1.25; Glu, 1.11; Phe, 0.84; Lys, <sup>a</sup> |
| Step 2                               | Ala, 0.00; Val, 1.88; Thr, 1.04; Asp, 1.07; Pro, 1.10; Glu, 1.06; Phe, 0.85; Lys, <sup>a</sup> |
| Step 3                               | Ala, 0.00; Val, 1.93; Thr, 0.25; Asp, 1.06; Pro, 0.89; Glu, 1.12; Phe, 1.00; Lys, <sup>a</sup> |
| Step 4                               | Ala, 0.00; Val, 1.86; Thr, 0.00; Asp, 0.47; Pro, 1.12; Glu, 1.13; Phe, 0.90; Lys, <sup>a</sup> |
| Step 5                               | Ala, 0.00; Val, 2.02; Thr, 0.00; Asp, 0.00; Pro, 0.37; Glu, 1.29; Phe, 0.69; Lys, <sup>a</sup> |
| Leucine amino peptidase A<br>30 min. | Ala, 1.00; Val, 0.98; Thr, 0.48; Asn, 0.42   |
| Subtilopeptidase peptides            |  |
| S1 ( $E_{Asp}$ , 0.35)               | Ala, 0.82; Val, 0.96; Thr, 0.98; Asp, 1.10; Pro, 1.10; Glu, 1.03                               |
| Edman degradation                    |  |
| Step 1                               | Ala, 0.00; Val, 0.82; Thr, 0.89; Asp, 1.06; Pro, 1.29; Glu, 0.95                               |
| Step 2                               | Ala, 0.00; Val, 0.00; Thr, 0.87; Asp, 1.09; Pro, 0.98; Glu, 1.06                               |
| Step 3                               | — Val, 0.00; Thr, 0.00; Asp, 0.93; Pro, 1.11; Glu, 0.97  |
| Step 4                               | — — Thr, 0.00; Asp, 0.60; Pro, 0.98; Glu, 1.02   |
| Step 5                               | — — — Asp, 0.00; Pro, 0.50; Glu, 1.00  |
| S2 (neutral)                         | Phe, 1.00  |
| S3 (neutral)                         | Val, (2)   |
| S4 ( $E_{Lys}$ , 0.51)               | Val, 1.03; Lys, 0.97   |
| Thermolytic peptides                 |  |
| Th1 ( $E_{Asp}$ , 0.32)              | Ala, 0.96; Val, 1.04; Thr, 1.00; Asp, 0.99; Pro, 1.10; Glu, 1.03; Phe, 0.88                    |
| Th2 ( $E_{Lys}$ , 0.47)              | Val, 2.18; Lys, 1.00   |
| Th3 ( $E_{Lys}$ , 0.40)              | Phe, 0.93; Val, 2.07; Lys, 1.00  |

<sup>a</sup> Not determined.

These were separated by electrophoresis at pH 6.5 and 1.9 and finally by paper chromatography with the butanol-acetic acid-water system. The amino acid sequence of S1 was derived from five steps of Edman degradation. The acidic electrophoretic mobility at pH 6.5 suggests that glutamic acid must be in the acid form. Since the NH<sub>2</sub>-terminal sequence of  $\alpha$ T7 was confirmed by the Edman degradation on S1, this is placed in the NH<sub>2</sub>-terminus of  $\alpha$ T7. Since lysine is placed at the COOH-terminus of  $\alpha$ T7 and considering the specificity of trypsin, the structure of S4 can only be *Val-Lys*. S2 is found to be free phenylalanine and S3 a dipeptide *Val-Val*. For confirmation and completion of the sequence,  $\alpha$ T7 was digested with thermolysin. Three main subpeptides were formed, Th1, Th2, and Th3, which were isolated by electrophoresis at pH 6.5. Th1 gives the necessary overlap for S1 and S2, Th2 overlaps S3 and S4 and Th3 overlaps S2 and S3.

Table 5. Amino acid sequences of peptides  $\alpha$ T8,  $\alpha$ T9,  $\alpha$ T10, and  $\alpha$ T12.

|                                |   |      |      |      |      |                   |
|--------------------------------|---|------|------|------|------|-------------------|
| Sequence $\alpha$ T8           | Ala - Ala - Val - Gly - Asp - Lys                     |      |      |      |      |                   |
| Edman degradation              |   |      |      |      |      |                   |
| Step 1                         | 0.00  | 1.15 | 0.91 | 0.96 | 0.98 | <sup>a</sup>      |
| Step 2                         | 0.00  | 0.35 | 0.90 | 1.04 | 1.06 | <sup>a</sup>      |
| Step 3                         | —   | 0.00 | 0.00 | 0.99 | 1.01 | <sup>a</sup>      |
| Step 4                         | —   | —    | 0.00 | 0.64 | 1.00 | <sup>a</sup>      |
| Sequence $\alpha$ T9           | Asp - Ser - Ala - Gly - Gln - Leu - Lys               |      |      |      |      |                   |
| Edman degradation              |   |      |      |      |      |                   |
| Step 1                         | 0.00  | 0.99 | 1.03 | 1.00 | 1.00 | 0.96 <sup>a</sup> |
| Step 2                         | 0.00  | 0.00 | 1.03 | 1.03 | 1.05 | 0.89 <sup>a</sup> |
| Step 3                         | —   | 0.00 | 0.30 | 1.09 | 1.03 | 0.88 <sup>a</sup> |
| Step 4                         | —   | —    | 0.00 | 0.70 | 1.02 | 0.98 <sup>a</sup> |
| Thermolytic peptides           |   |      |      |      |      |                   |
| Th1 ( $E_{\text{Asp}}$ , 0.48) | Asp, 0.99; Ser, 0.90; Ala, 1.10; Gly, 1.02; Gln, 1.01 |      |      |      |      |                   |
| Th2                            | Leu, 0.96; Lys, 1.04                                  |      |      |      |      |                   |
| Sequence $\alpha$ T10          | Glu - Ala - Leu - Leu - Lys                           |      |      |      |      |                   |
| Edman degradation              |   |      |      |      |      |                   |
| Step 1                         | 0.00  | 0.98 | 2.01 |      |      | <sup>a</sup>      |
| Step 2                         | 0.00  | 0.00 | 2.00 |      |      | <sup>a</sup>      |
| Sequence $\alpha$ T12          | Val - Ala - Phe - Thr - Glu - Lys                     |      |      |      |      |                   |
| Edman degradation              |   |      |      |      |      |                   |
| Step 1                         | 0.00  | 0.96 | 1.01 | 0.97 | 1.07 | <sup>a</sup>      |
| Step 2                         | 0.00  | 0.00 | 0.88 | 1.06 | 1.06 | <sup>a</sup>      |
| Step 3                         | —   | 0.00 | 0.00 | 0.93 | 1.07 | <sup>a</sup>      |
| Step 4                         | —   | —    | 0.00 | 0.00 | 1.00 | <sup>a</sup>      |

<sup>a</sup> Not determined.

*Peptide aT8.* The results of the Edman degradation presented in Table 5 permit the determination of the amino acid sequence of this peptide. The aspartic acid residue in the peptide must be in the acid form because of the neutral mobility of the peptide on electrophoresis at pH 6.5.

*Peptide aT9.* The Edman degradation of this peptide was carried out through four steps which yielded the NH<sub>2</sub>-terminal sequence *Asp-Ser-Ala-Gly-* (Table 5). Digestion of *aT9* with thermolysin yielded two peptides, Th1 and Th2, which were purified by paper electrophoresis at pH 6.5. Th1 was an acidic pentapeptide which, after removal of aspartic acid by the Edman technique, yielded an electrophoretically neutral peptide indicating the presence of glutamine. Th2 was a basic peptide containing leucine and lysine. Since lysine is the COOH-terminus of *aT9*, the sequence of Th2 is *Leu-Lys*.

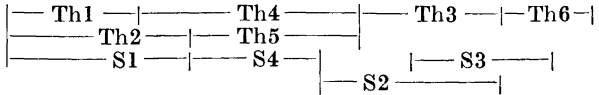
*Peptide aT10.* Two steps of Edman degradation (Table 5) established the complete sequence of this peptide. The glutamic acid residue in the peptide was shown to be in the acid form by its neutral behaviour on paper electrophoresis at pH 6.5.

*Peptide aT11.* No sequence studies were performed as the structure of this histidine peptide has been reported earlier.<sup>18</sup>

*Peptide aT12.* Four steps of Edman degradation established the sequence of this peptide (Table 5). The glutamic acid residue must be in the acid form because of the neutral electrophoretic mobility of the whole peptide at pH 6.5. The sequence of the first four amino acids of this peptide is identical to that obtained for the NH<sub>2</sub>-terminal end of the intact peptide chain of *Lba*, which indicates that *aT12* represents the NH<sub>2</sub>-terminal sequence of *Lba*.

*Peptide aT13.* One step of Edman degradation gave no change in the amino acid composition of the peptide compared to the untreated sample. It was assumed that a pyrrolidone carboxylic acid occupied the NH<sub>2</sub>-terminal position and that the transformation of a glutamine residue may have occurred during isolation of the peptide. When the peptide was subjected to thermolytic hydrolysis five principal peptides were formed, Th1, Th2, Th3, Th4, and Th5. These were separated electrophoretically at pH 6.5 and further at pH 1.9. Table 6 shows their amino acid composition. Since Th1, like *aT13* itself, shows no NH<sub>2</sub>-terminal amino acid by Edman degradation, Th1 had to represent the NH<sub>2</sub>-terminal position of *aT13*. The amino acid sequence of Th1 was derived after partial hydrolysis of the peptide with 12 N HCl at 37°C. The hydrolysate was resolved by electrophoresis at pH 6.5, and free glutamic acid, aspartic acid, and alanine were isolated in addition to a dipeptide Th1-H1 the sequence of which was *Asp-Ala* as shown by one step of Edman degradation (Table 6). The aspartic acid residue of Th1 is present in the acid form. When precautions were taken to avoid degradation of the assumed labile NH<sub>2</sub>-terminal glutamine, a thermolytic peptide Th1A was isolated containing aspartic and glutamic acid and alanine. Its electrophoretic mobility shows only one acidic group to be free, which indicates that glutamic acid is amidated. Two steps of dansyl-Edman degradation yielded the sequence *Gln-Asp-Ala*. The composition of Th2 suggests its structure to be identical with that of Th1 elongated by one amino acid, leucine. This conclusion is supported by carboxypeptidase A, establishing Leu as the COOH-terminus of the peptide. Its electrophoretic mobility showed two

Table 6. Amino acid sequence of peptide  $\alpha$ T13.

|   |   |
|---|---|
| Sequence <sup>b</sup>                               | Glp-Asp-Ala-Leu-Val-Ser-Ser-Ser-Phe-Glu-Ala-Phe-Lys<br> |
| Edman degradation<br>Step 1                         | Glu, 1.95; Asp, 1.09; Ala, 2.09; Leu, 1.06; Val, 1.03;<br>Ser, 2.86; Phe, 1.77; Lys <sup>a</sup>  |
| Thermolytic peptides<br>Th 1 ( $E_{Asp}$ , 1.07)    | Asp, 1.03; Glu, 0.96; Ala, 1.00   |
| Hydrolysis in 12 N HCl<br>Th1-H1                    | Asp, 1.00; Ala, 1.00  |
| Edman degradation<br>Step 1                         | Asp, 0.00; Ala, 1.00  |
| Th1A ( $E_{Asp}$ , 0.59)                            | Asp, 1.04; Glu, 0.99; Ala, 0.98   |
| Edman degradation<br>Step 1                         | Direct examination of dansyl derivatives  |
| Step 2  | DNS-Glu<br>DNS-Asp  |
| Th2 ( $E_{Asp}$ , 0.92)                             | Asp, 0.99; Glu, 0.85; Ala, 0.99; Leu, 1.16  |
| Carboxypeptidase A 100 min                          | Leu, 0.52   |
| Th2A ( $E_{Asp}$ , 0.50)                            | Amino acid composition identical with that of Th2   |
| Th3 ( $E_{Asp}$ , 0.52)                             | Phe, 0.75; Glu, 1.00; Ala, 1.00   |
| Edman degradation<br>Step 1                         | Phe, 0.00; Glu, 1.00; Ala, 1.00   |
| Step 2  | — Glu, 0.00; Ala, 1.00  |
| Th4 (neutral)                                       | Leu, 0.91; Val, 1.01; Ser, 3.13   |
| Edman degradation<br>Step 1                         | Leu, 0.00; Val, 0.94; Ser, 3.05   |
| Step 2  | — Val, 0.00; Ser, 3.00  |
| Th5 (neutral)                                       | Val, 1.13; Ser, 2.86  |
| Th6 ( $E_{Lys}$ , 0.63)                             | Phe, 1.00; Lys, 1.00  |
| Subtilopeptidase peptides<br>S1 ( $E_{Asp}$ , 0.92) | Glu, 0.95; Asp, 0.98; Ala, 1.00; Leu, 1.07  |
| S2 ( $E_{Asp}$ , 0.52)                              | Glu, 1.00; Ala, 1.05; Phe, 0.95   |
| Edman degradation<br>Step 1                         | Glu, 0.00; Ala, 1.01; Phe, 0.99   |
| Step 2  | — Ala, 0.00; Phe, 1.00  |
| S3 ( $E_{Asp}$ , 0.43)                              | Ser, 1.10; Phe, 1.02; Glu, 1.02; Ala, 0.86  |
| S4  | Val, 0.86; Ser, 2.15  |
| Edman degradation<br>Step 1                         | Val, 0.00; Ser, 2.00  |

<sup>a</sup> Not determined. <sup>b</sup> Glp, abbreviation used for pyrrolidone carboxylic acid.



acidic groups to be free. The composition of Th2A is identical to that of Th2 except that it has only one free acidic group, a relationship identical to that of Th1 and Th1A. Th6 contains the lysine residue of *aT13*, and therefore occupies the COOH-terminal portion of the peptide. Its sequence, *Phe-Lys*, is deduced by considering the specificity of trypsin. Th3 was determined as *Phe-Glu-Ala* by two steps of Edman degradation. The glutamic acid residue of this peptide must be in the acid form because of the acidic electrophoretic mobility of the peptide at pH 6.5. Th4 was deduced to be *Leu-Val-Ser-Ser-Ser* by two steps of Edman degradation. Th5 was assumed to be *Val-Ser-Ser-Ser* by referring to the sequence of Th4. Th3 and Th4 both represent the middle portion of the peptide *aT13*, Th2 giving the overlap between Th1 and Th4. The order of these components was further established and confirmed by the overlapping peptides obtained from a subtilo-peptidase A digest of *aT13*. Four peptides, S1, S2, S3, and S4, were isolated by electrophoresis at pH 6.5 and 1.9. The amino acid composition of these subpeptides is given in Table 6. S1 contained the NH<sub>2</sub>-terminal four residues and electrophoresis at pH 6.5 showed it to have two free acidic groups. Two steps of Edman degradation and acidic mobility at pH 6.5 showed the sequence of S2 to be *Glu-Ala-Phe*. The sequence of S4 was deduced to be *Val-Ser-Ser* from one step of Edman degradation. S1 gives the overlap of Th1 and Th4, S2 that of Th4 and Th3, and S3 that of Th3 and Th6.

*Peptide aT14.* The NH<sub>2</sub>-terminal sequence of the peptide was found to be *Asp-Leu-Phe-Ser-Phe-Leu-* as revealed by six steps of subtractive and dansyl Edman procedures (Table 7). Hydrolysis in 0.03 N HCl produced free aspartic acid and four peptide fragments, A1, A2, A3, and A4 which were isolated by paper electrophoresis at pH 6.5 and further at pH 1.9. A1 represents the NH<sub>2</sub>-terminal part of the peptide containing five of the NH<sub>2</sub>-terminal amino acid residues. Its sequence is therefore concluded to be *Leu-Phe-Ser-Phe-Leu-Ala*. A4 represents the COOH-terminal portion of the peptide. Its sequence had to be *Pro-Lys* from the specificity of trypsin; the sequence was also confirmed by dansylation (Table 7). A2 and A3 are derived from the middle portion of the original peptide and the sequences were found to be *Pro-Thr* and *Gly-Val*, respectively (Table 7). The liberation of *Pro-Thr*, *Gly-Val*, and *Pro-Lys* indicates the sequences *Asx-Pro-Thr-Asx*, *Asx-Gly-Val-Asx*, and *Asx-Pro-Lys* considering the specificity of the mild acid hydrolysis. For completion of the sequence, *aT14* was subjected to hydrolysis with subtilo-peptidase A. In addition to free serine, leucine, and lysine, six peptide fragments were formed, S1, S2, S3, S4, S7, and S8, which were isolated by electrophoresis at pH 6.5 and 1.9. Table 7 shows their amino acid composition. S1 and S2 both represent the NH<sub>2</sub>-terminal portion of the original peptide, *aT14*. The electrophoretic mobility of these two peptides at pH 6.5 indicated the presence of nonamidated aspartic acid. The sequence for S1 was found to be *Asp-Leu* and for S2 *Asp-Leu-Phe* by dansyl-Edman degradation. S7 and S8 both contained lysine and therefore represent the COOH terminus of the original peptide. S8 is a basic peptide on electrophoresis at pH 6.5, indicating the fourth aspartic acid residue of the original peptide to be amidated. Three steps of dansyl-Edman showed its sequence to be *Gly-Val-Asn-Pro-Lys*. S7, which contains the third and fourth aspartic

Table 7. Amino acid sequence of peptide  $\alpha$ T14.

| Sequence                   | Asp-Leu-Phe-Ser-Phe-Leu-Ala-Asn-Pro-Thr-Asp-Gly-Val-Asn-Pro-Lys   |
|----------------------------|---|
|                            |   |
| Edman degradation          |   |
| Step 1                     | Asp, 2.97; Leu, 1.87; Phe, 1.67; Ser, 1.00; Ala, 1.04; Pro, 2.23; Thr, 1.02; Gly, 1.06; Val, 0.93; Lys <sup>a</sup> |
| Step 2                     | Asp, 3.07; Leu, 0.98; Phe, 1.62; Ser, 1.22; Ala, 1.02; Pro, 2.08; Thr, 1.01; Gly, 1.11; Val, 0.90; Lys <sup>a</sup> |
| Step 3                     | Asp, 2.93; Leu, 0.95; Phe, 0.87; Ser, 1.20; Ala, 1.03; Pro, 2.27; Thr, 0.90; Gly, 1.05; Val, 0.83; Lys <sup>a</sup> |
| Step 4                     | Direct examination of dansyl derivatives  |
| Step 5                     | DNS-Ser   |
| Step 6                     | DNS-Phe   |
|                            | DNS-Leu   |
| Hydrolysis with 0.03 N HCl |   |
| A1                         | Leu, 2.04; Phe, 1.84; Ser, 1.15; Ala, 0.98  |
| A2                         | Pro, 1.15; Thr, 0.85  |
| Dansylation                | DNS-Pro   |
| A3                         | Gly, 1.01; Val, 0.99  |
| Dansylation                | DNS-Gly   |
| A4                         | Pro, 0.98; Lys, 1.02  |
| Dansylation                | DNS-Pro   |
| Subtilopeptidase peptides  |   |
| S1 ( $E_{Asp}$ , 0.67)     | Asp, 1.03; Leu, 0.97  |
| Dansylation                | DNS-Asp   |
| S2 ( $E_{Asp}$ , 0.51)     | Asp, 1.05; Leu, 0.98; Phe, 0.98   |
| Edman degradation          |   |
| Step 1                     | Asp, 0.25; Leu, 1.0; Phe, 0.85  |
| Step 2                     | — Leu, 0.61; Phe, 1.00  |
| S3 ( $E_{Asp}$ , 0.31)     | Leu, 0.87; Ala, 1.11; Asp, 1.93; Pro, 1.05; Thr, 0.89; Gly, 1.14  |
| Dansylation                | DNS-Leu   |
| S4 (neutral)               | Leu, 0.93; Ala, 0.93; Asp, 1.13   |
| Edman degradation          |   |
| Step 1                     | Leu, 0.00; Ala, 0.96; Asp, 1.04   |
| Step 2                     | — Ala, 0.00; Asp, 1.00  |
| S5                         | Ser   |
| S6                         | Leu   |
| S7 (neutral)               | Thr, 1.01; Asp, 2.00; Gly, 1.20; Val, 0.88; Pro, 1.20; Lys, 0.98  |
| Edman degradation          | Direct examination of dansyl derivatives  |
| Step 1                     | DNS-Thr   |
| Step 2                     | DNS-Asp   |
| Step 3                     | DNS-Gly   |
| S8 ( $E_{Lys}$ , 0.51)     | Gly, 1.03; Val, 1.00; Asp, 1.33; Pro, 0.60; Lys, 0.97   |
| Edman degradation          | Direct examination of dansyl derivatives  |
| Step 1                     | DNS-Gly   |
| Step 2                     | DNS-Val   |
| Step 3                     | DNS-Asp   |
| S9                         | Lys   |

<sup>a</sup> Not determined.

acid residue of  $\alpha$ T14, was found to be neutral on electrophoresis at pH 6.5, so indicating the third aspartic acid residue to be in the acid form. After three steps of dansyl-Edman degradation and by referring to the sequence of S8, the sequence of S7 is concluded to be *Thr-Asp-Gly-Val-Asn-Pro-Lys*. S3 and S4 represent the middle portion of the original peptide. The neutral electrophoretic mobility of S4 at pH 6.5 showed the second aspartic acid residue of  $\alpha$ T14 to be amidated, the sequence of the peptide being *Leu-Ala-Asn* as shown by two steps of Edman degradation. S3 is an acidic peptide which contains the second and the third aspartic acid residue of  $\alpha$ T14 and confirms that third aspartic acid residue is in the acid form. The sequence of S3 is concluded to be *Leu-Ala-Asn-Pro-Thr-Asp-Gly* referring to the sequences of A2, S4, and S7.

*Peptide  $\alpha$ T15.* The amino acid composition of this peptide shows it to be identical to that of  $\alpha$ T13 except for two phenylalanyl residues which are assumed to be exchanged into two tyrosyl residues. Subpeptides of a subtiloypeptidase digest were studied to confirm this assumption. Six peptides could be isolated electrophoretically at pH 6.5 and 1.9: S1, S2, S3, S4, S5, and S6 (Table 8). Two steps of Edman degradation and an acidic mobility on electrophoresis at pH 6.5 showed the sequence of S1 to be *Glu-Ala-Tyr*.

Table 8. Amino acid sequence of peptide  $\alpha$ T15.

| Sequence               | [Glu-Asn-Ala-Leu]-[Val-Ser-Ser]-[Ser, Tyr]-[Glu-Ala-Tyr]-Lys<br> ----- S2 -----   ---- S4 ----   ----- S1 ----- S6- |
|------------------------|---|
| Subtilisin peptides    |   |
| S1 ( $E_{Asp}$ , 0.45) | Glu, 1.05; Ala, 1.04; Tyr, 0.91   |
| Edman degradation      |   |
| Step 1                 | <i>Glu</i> , 0.00; Ala, 1.13; Tyr, 0.87   |
| Step 2                 | — <i>Ala</i> , 0.65; Tyr, 1.00  |
| S2 ( $E_{Asp}$ , 0.42) | Glu, 1.01; Asp, 1.04; Ala, 1.05; Leu, 0.90  |
| Edman degradation      |   |
| Step 1                 | <i>Glu</i> , 0.00; Asp, 1.07; Ala, 1.05; Leu, 0.88  |
| Step 2                 | — <i>Asp</i> , 0.48; Ala, 1.17; Leu, 0.83   |
| Step 3                 | —     — <i>Ala</i> , 0.37; Leu, 0.83  |
| Carboxypeptidase A     |   |
| 30 min                 | Leu, 1.00; Ala, 0.39; Asn (Ser) 0.10  |
| S3                     | Ser, 1.00 (2)   |
| S4                     | Val, 1.03; Ser, 1.97  |
| Edman degradation      |   |
| Step 1                 | <i>Val</i> , 0.00; Ser, 2.00  |
| S5                     | Val, 0.92; Ser, 1.13  |
| Edman degradation      |   |
| Step 1                 | <i>Val</i> , 0.00; Ser, 1.00  |
| S6 ( $E_{Lys}$ , 1.00) | Lys, 1.00   |

Three steps of Edman degradation and carboxypeptidase A digestion indicated the sequence of S2 to be *Glu-Asn-Ala-Leu*. On the basis of electrophoresis at pH 1.9, S3 seems to be *Ser-Ser*. S4 was shown to be *Val-Ser-Ser* and S5 *Val-Ser*, both by one step of Edman degradation. S6 was found to be free lysine. The sequences of the subpeptides confirm our assumption of a homology between *aT13* and *aT15* and are put in order according to the sequence of *aT13*.

*Peptide aT16*. Five steps of Edman degradation established the sequence of this peptide (Table 9). The aspartic and glutamic acid residues must be

Table 9. Amino acid sequence of peptide *aT16*.

|                                       |  |
|---------------------------------------|--|
| Sequence                              | Trp-Ser-Asp-Glu-Leu-Ser-Arg  |
|                                       | ——— Th1 ——— ——— Th2 ———  |
| Edman degradation                     |  |
| Step 1                                | <i>Trp</i> , 0.00; <i>Ser</i> , 1.77; <i>Asp</i> , 1.13; <i>Glu</i> , 1.15; <i>Leu</i> , 0.95; <i>Arg</i> <sup>a</sup> |
| Step 2                                | — <i>Ser</i> , 1.07; <i>Asp</i> , 1.03; <i>Glu</i> , 1.06; <i>Leu</i> , 0.88; <i>Arg</i> <sup>a</sup>                  |
| Step 3                                | — <i>Ser</i> , 1.01; <i>Asp</i> , 0.49; <i>Glu</i> , 1.06; <i>Leu</i> , 0.93; <i>Arg</i> <sup>a</sup>                  |
| Step 4                                | — <i>Ser</i> , 1.06; <i>Asp</i> , 0.00; <i>Glu</i> , 0.63; <i>Leu</i> , 0.94; <i>Arg</i> <sup>a</sup>                  |
| Step 5                                | — <i>Ser</i> , 1.00; <i>Asp</i> , 0.00; <i>Glu</i> , 0.56; <i>Leu</i> , 0.41; <i>Arg</i> <sup>a</sup>                  |
| Thermolysin peptides                  |  |
| Th1 ( <i>E</i> <sub>Asp</sub> , 0.64) | <i>Trp</i> , +; <i>Ser</i> , 0.92; <i>Asp</i> , 1.02; <i>Glu</i> , 1.06  |
| Th2 ( <i>E</i> <sub>Lys</sub> , 0.53) | <i>Leu</i> , 0.96; <i>Ser</i> , 1.09; <i>Arg</i> , 0.95  |
| Edman degradation                     |  |
| Step 1                                | <i>Leu</i> , 0.00; <i>Ser</i> , 1.00; <i>Arg</i> , <sup>a</sup>  |

<sup>a</sup> Not determined.

in the acid form because of the electrophoretic mobility of the whole peptide at pH 6.5. The sequence was further confirmed by thermolytic fragments, Th1 and Th2, which were isolated by paper electrophoresis at pH 6.5. Th1 contained the four NH<sub>2</sub>-terminal residues and Th2 represented the COOH-terminal position of the whole peptide. One step of Edman degradation on Th2 confirmed the sequence *Leu-Ser-Arg*.

*Peptide aT17*. The Edman degradation revealed the NH<sub>2</sub>-terminal alanine of this peptide (Table 10). Digestion by chymotrypsin yielded two principal components, C1 and C2, which were separated by electrophoresis at pH 6.5. C1 represents the NH<sub>2</sub>-terminal portion of the whole peptide because it contains the NH<sub>2</sub>-terminal alanine. C2 represents the COOH-terminal portion of the whole peptide because it contains lysine. Three steps of Edman degradation yielded the COOH-terminal sequence *Ala-Ala-Ala-Ile-Lys*. On hydrolysis with thermolysin, five peptides were formed, Th1, Th2, Th3, Th4, and Th5, which were separated on electrophoresis at pH 6.5 and 1.9. Th3 represented the NH<sub>2</sub>-terminal portion of the whole peptide. Two steps of Edman degradation established the sequence *Ala-Trp-Glu*. Th5 contains the lysine residue of peptide *aT17* and therefore represents the COOH-terminal

Table 10. Amino acid sequence of peptide  $\alpha T17$ .

|                         |  |
|-------------------------|--|
| Sequence                | Ala-Trp-Glu-Val-Ala-Tyr-Asp-Glu-Leu-Ala-Ala-Ala-Ile-Lys  |
|                         |  |
| Edman degradation       |  |
| Step 1                  | Ala, 4.02; Trp, 1.00; Glu, 2.04; Val, 0.88; Tyr, 0.92; Asp, 1.25; Leu, 1.25; Ile, 0.80; Lys <sup>a</sup> |
| Chymotryptic peptides   |  |
| C1 ( $E_{Asp}$ , 0.47)  | Ala, 1.92; Trp, 1.00; Glu, 1.99; Val, 0.97; Tyr, 0.43; Asp, 1.12; Leu, 1.00                              |
| Dansylation             | DNS-Ala  |
| C2 ( $E_{Lys}$ , 0.48)  | Ala, 3.09; Ile, 0.92; Lys, 0.99  |
| Edman degradation       |  |
| Step 1                  | Ala, 2.16; Ile, 0.84; Lys <sup>a</sup>   |
| Step 2                  | Ala, 1.17; Ile, 0.83; Lys <sup>a</sup>   |
| Step 3                  | Ala, 0.49; Ile, 0.84; Lys <sup>a</sup>   |
| Thermolytic peptides    |  |
| Th1 ( $E_{Asp}$ , 0.65) | Val, 0.98; Ala, 1.03; Tyr, 0.90; Asp, 1.13; Glu, 0.99  |
| Edman degradation       |  |
| Step 1                  | Val, 0.00; Ala, 0.93; Tyr, 1.03; Asp, 1.03; Glu, 1.01  |
| Step 2                  | Val, 0.00; Ala, 0.28; Tyr, 0.46; Asp, 1.14; Glu, 1.00  |
| Step 3                  | — Ala, 0.00; Tyr, 0.29; Asp, 1.06; Glu, 1.00   |
| Step 4                  | — — Tyr, 0.00; Asp, 0.76; Glu, 1.00  |
| Th2 ( $E_{Asp}$ , 0.59) | Val, 0.91; Ala, 3.19; Tyr, 0.53; Asp, 1.00; Glu, 0.92; Leu, 0.96   |
| Th3 ( $E_{Asp}$ , 0.41) | Ala, 0.98; Trp, +; Glu, 1.02   |
| Edman degradation       |  |
| Step 1                  | Ala, 0.26; Trp, +; Glu, 1.00   |
| Step 2                  | Ala, 0.00; Trp, —; Glu, 1.00   |
| Th4 (neutral)           | Leu, 0.87; Ala, 3.13   |
| Edman degradation       |  |
| Step 1                  | Leu, 0.30; Ala, 3.00   |
| Th5 ( $E_{Lys}$ , 0.72) | Ile, 0.99; Lys, 1.01   |

<sup>a</sup> Not determined.

portion of the peptide. Its sequence had to be *Ile-Lys*. Th1 is a component derived from the middle portion of the whole peptide and was determined to be *Val-Ala-Tyr-Asp-Glu* by four steps of Edman degradation. The electrophoretic mobility of the peptide at pH 6.5 shows aspartic and glutamic acid to be in the acid form. Th4 is another component from the inner portion of the whole peptide. One step of Edman degradation yielded the sequence *Leu-Ala-Ala-Ala*. Th2 gives the overlap between Th1 and Th4.

*Peptide*  $\alpha T18$ . Alanine was identified as the  $NH_2$ -terminal residue by dansylation. Thermolytic digestion of the peptide yielded eight peptide frag-

Table 11. Amino acid sequence of peptide  $\alpha$ T18.

|                          |   |     |     |     |  |     |     |     |  |  |     |  |  |    |    |  |    |  |    |    |    |
|--------------------------|---|-----|-----|-----|--|-----|-----|-----|--|--|-----|--|--|----|----|--|----|--|----|----|----|
| Sequence                 | Ala-Asn-Ile-Pro-Gln-Tyr-Ser-Val-Val-Phe-Tyr-Thr-Ser-Ile-Leu-Glu-Lys   |     |     |     |  |     |     |     |  |  |     |  |  |    |    |  |    |  |    |    |    |
|                          | <table border="1" style="width: 100%; border-collapse: collapse; text-align: center;"> <tr> <td style="width: 25%;">Th1</td> <td style="width: 25%;">Th4</td> <td style="width: 25%;">Th8</td> <td style="width: 25%;"></td> </tr> <tr> <td>Th2</td> <td>Th3</td> <td>Th7</td> <td></td> </tr> <tr> <td></td> <td>Th5</td> <td></td> <td></td> </tr> <tr> <td>S1</td> <td>S4</td> <td></td> <td>S6</td> </tr> <tr> <td></td> <td>S2</td> <td>S3</td> <td>S5</td> </tr> </table> | Th1 | Th4 | Th8 |  | Th2 | Th3 | Th7 |  |  | Th5 |  |  | S1 | S4 |  | S6 |  | S2 | S3 | S5 |
| Th1                      | Th4   | Th8 |     |     |  |     |     |     |  |  |     |  |  |    |    |  |    |  |    |    |    |
| Th2                      | Th3   | Th7 |     |     |  |     |     |     |  |  |     |  |  |    |    |  |    |  |    |    |    |
|                          | Th5   |     |     |     |  |     |     |     |  |  |     |  |  |    |    |  |    |  |    |    |    |
| S1                       | S4  |     | S6  |     |  |     |     |     |  |  |     |  |  |    |    |  |    |  |    |    |    |
|                          | S2  | S3  | S5  |     |  |     |     |     |  |  |     |  |  |    |    |  |    |  |    |    |    |
| Dansylation              | DNS-Ala   |     |     |     |  |     |     |     |  |  |     |  |  |    |    |  |    |  |    |    |    |
| Thermolysin peptides     |   |     |     |     |  |     |     |     |  |  |     |  |  |    |    |  |    |  |    |    |    |
| Th1 (neutral)            | Ala, 0.98; Asp, 1.00; Ile, 1.15; Pro, 1.25; Glu, 1.09; Tyr, 0.87; Ser, 0.91   |     |     |     |  |     |     |     |  |  |     |  |  |    |    |  |    |  |    |    |    |
| Hydrazinolysis           | Ser, 0.80   |     |     |     |  |     |     |     |  |  |     |  |  |    |    |  |    |  |    |    |    |
| Th2 (neutral)            | Ala, 0.86; Asp, 1.03; Ile, 1.04; Pro, 1.44; Glu, 1.07   |     |     |     |  |     |     |     |  |  |     |  |  |    |    |  |    |  |    |    |    |
| Edman degradation        |   |     |     |     |  |     |     |     |  |  |     |  |  |    |    |  |    |  |    |    |    |
| Step 1                   | <i>Ala</i> , 0.00; Asp, 0.98; Ile, 1.04; Pro, 0.84; Glu, 1.13   |     |     |     |  |     |     |     |  |  |     |  |  |    |    |  |    |  |    |    |    |
| Step 2                   | — <i>Asp</i> , 0.12; Ile, 1.01; Pro, 0.88; Glu, 1.12  |     |     |     |  |     |     |     |  |  |     |  |  |    |    |  |    |  |    |    |    |
| Step 3                   | — — <i>Ile</i> , 0.11; Pro, 0.91; Glu, 1.09   |     |     |     |  |     |     |     |  |  |     |  |  |    |    |  |    |  |    |    |    |
| Step 4                   | — — — <i>Pro</i> , 0.00; Glu, 1.00  |     |     |     |  |     |     |     |  |  |     |  |  |    |    |  |    |  |    |    |    |
| Th3 (neutral)            | Val, 1.90; Phe, 1.09; Tyr, 0.60   |     |     |     |  |     |     |     |  |  |     |  |  |    |    |  |    |  |    |    |    |
| (70 h hydrolysis)        |   |     |     |     |  |     |     |     |  |  |     |  |  |    |    |  |    |  |    |    |    |
| Edman degradation        | Direct examination of dansyl derivatives  |     |     |     |  |     |     |     |  |  |     |  |  |    |    |  |    |  |    |    |    |
| Step 1                   | DNS-Val; DNS-Val-Val  |     |     |     |  |     |     |     |  |  |     |  |  |    |    |  |    |  |    |    |    |
| Step 2                   | DNS-Val   |     |     |     |  |     |     |     |  |  |     |  |  |    |    |  |    |  |    |    |    |
| Step 3                   | DNS-Phe   |     |     |     |  |     |     |     |  |  |     |  |  |    |    |  |    |  |    |    |    |
| Th4 (neutral)            | Val, 1.86; Phe, 1.23; Tyr, 0.60; Thr, 0.80; Ser, 0.92   |     |     |     |  |     |     |     |  |  |     |  |  |    |    |  |    |  |    |    |    |
| (70 h hydrolysis)        |   |     |     |     |  |     |     |     |  |  |     |  |  |    |    |  |    |  |    |    |    |
| Dansylation              | DNS-Val; DNS-Val-Val  |     |     |     |  |     |     |     |  |  |     |  |  |    |    |  |    |  |    |    |    |
| Hydrazinolysis           | Ser, 0.92   |     |     |     |  |     |     |     |  |  |     |  |  |    |    |  |    |  |    |    |    |
| Hydrolysis in 12 N HCl   |   |     |     |     |  |     |     |     |  |  |     |  |  |    |    |  |    |  |    |    |    |
| Th4-H1 (70 h hydrolysis) | Val, 2.02; Phe, 0.98  |     |     |     |  |     |     |     |  |  |     |  |  |    |    |  |    |  |    |    |    |
| Th4-H2                   | Tyr, 0.55; Thr, 1.04; Ser, 0.96   |     |     |     |  |     |     |     |  |  |     |  |  |    |    |  |    |  |    |    |    |
| Th5 (neutral)            | Tyr, 0.86; Thr, 1.09; Ser, 1.09   |     |     |     |  |     |     |     |  |  |     |  |  |    |    |  |    |  |    |    |    |
| Edman degradation        | Direct examination of dansyl derivatives  |     |     |     |  |     |     |     |  |  |     |  |  |    |    |  |    |  |    |    |    |
| Step 1                   | DNS-Tyr   |     |     |     |  |     |     |     |  |  |     |  |  |    |    |  |    |  |    |    |    |
| Step 2                   | DNS-Thr   |     |     |     |  |     |     |     |  |  |     |  |  |    |    |  |    |  |    |    |    |
| Th6                      | Tyr, 0.90; Ser, 1.00  |     |     |     |  |     |     |     |  |  |     |  |  |    |    |  |    |  |    |    |    |
| Edman degradation        |   |     |     |     |  |     |     |     |  |  |     |  |  |    |    |  |    |  |    |    |    |
| Step 1                   | <i>Tyr</i> , 0.00; Ser, 1.00  |     |     |     |  |     |     |     |  |  |     |  |  |    |    |  |    |  |    |    |    |
| Th7 (neutral)            | Thr, 0.87; Ser, 0.98; Ile, 0.89; Leu, 0.99; Glu, 1.11; Lys, 1.05  |     |     |     |  |     |     |     |  |  |     |  |  |    |    |  |    |  |    |    |    |
| Edman degradation        | Direct examination of dansyl derivatives  |     |     |     |  |     |     |     |  |  |     |  |  |    |    |  |    |  |    |    |    |
| Step 1                   | DNS-Thr   |     |     |     |  |     |     |     |  |  |     |  |  |    |    |  |    |  |    |    |    |
| Step 2                   | DNS-Ser   |     |     |     |  |     |     |     |  |  |     |  |  |    |    |  |    |  |    |    |    |
| Step 3                   | DNS-Ile   |     |     |     |  |     |     |     |  |  |     |  |  |    |    |  |    |  |    |    |    |
| Step 4                   | DNS-Leu   |     |     |     |  |     |     |     |  |  |     |  |  |    |    |  |    |  |    |    |    |
| Th8 (neutral)            | Ile, 0.91; Leu, 0.98; Glu, 1.12; Lys, 0.99  |     |     |     |  |     |     |     |  |  |     |  |  |    |    |  |    |  |    |    |    |
| Edman degradation        |   |     |     |     |  |     |     |     |  |  |     |  |  |    |    |  |    |  |    |    |    |
| Step 1                   | <i>Ile</i> , 0.00; Leu, 0.87; Glu, 1.17; Lys <sup>a</sup>   |     |     |     |  |     |     |     |  |  |     |  |  |    |    |  |    |  |    |    |    |
| Step 2                   | — <i>Leu</i> , 0.00; Glu, 1.00; Lys <sup>a</sup>  |     |     |     |  |     |     |     |  |  |     |  |  |    |    |  |    |  |    |    |    |

Table 11. Continued.

|                           |  |
|---------------------------|--|
| Subtilopeptidase peptides |  |
| S1 (neutral)              | Ala, 0.80; Asp, 0.95; Ile, 1.12; Pro, 0.90; Glu, 1.16; Tyr, 1.07   |
| S2 (neutral)              | Ala, 0.97; Asp, 1.12; Ile, 1.10; Pro, 1.02; Glu, 1.03; Tyr, 1.22 (2); Ser, 1.05; Val, 0.68 (2) <sup>b</sup> ; Phe, 0.89; Thr, 0.81 |
| S3                        | Ser  |
| S4                        | Ser, 1.03; Val, 0.97   |
| Dansylation               | DNS-Ser  |
| S5 (neutral)              | Ile, 1.00; Leu, 1.00; Glu, 1.30; Lys, 1.17   |
| S6 (neutral)              | Glu, 0.92; Lys, 1.08   |

<sup>a</sup> Not determined. <sup>b</sup> Incomplete hydrolysis of the Val-Val bond in 18 h hydrolysis.

ments, Th1, Th2, Th3, Th4, Th5, Th6, Th7, and Th8, which were electrophoretically separated at pH 6.5 and 1.9. Th1 and Th2 are the only peptides containing the NH<sub>2</sub>-terminal alanine. Four steps of Edman degradation of Th2 established the sequence *Ala-Asn-Ile-Pro-Gln*. Since Th1 and Th2 were neutral on electrophoresis at pH 6.5, the glutamic and aspartic acid residues must be in amide form. Th6 was found to be *Tyr-Ser* after one step of Edman degradation. The COOH-terminal amino acid residue in Th1 was found to be serine by hydrazinolysis and the sequence of the peptide was concluded to be *Ala-Asn-Ile-Pro-Gln-Tyr-Ser* referring to the sequence of Th2 and Th6. Th8 represents the COOH-terminal portion of the peptide containing lysine. Two steps of Edman degradation established its sequence as *Ile-Leu-Glu-Lys*. Since the peptide was neutral on electrophoresis at pH 6.5, the glutamic acid residue is in the acid form. Th3, Th4, and Th5 represent the middle portion of peptide aT18. The presence of two valine residues in Th3 was confirmed by a 70 h acid hydrolysis of the peptide, which also indicated the sequence *Val-Val*, known to be highly resistant to acid hydrolysis. The dansyl-Edman procedure established the sequence *Val-Val-Phe-Tyr* for Th3. Two steps of dansyl-Edman degradation showed the sequence of Th5 to be *Tyr-Thr-Ser*. Dansylation of Th4 revealed *Val-Val* as the NH<sub>2</sub>-terminal sequence and hydrazinolysis identified serine as the COOH-terminal residue. Hydrolysis with 12 N HCl results in two main peptide fragments which were concluded to be *Val-Val-Phe* and *Tyr-Thr-Ser*. The sequence of Th4 is accordingly *Val-Val-Phe-Tyr-Thr-Ser*. The order of the thermolytic subpeptides was further established and their sequences were confirmed by overlapping peptides obtained from a subtilopeptidase A hydrolysis of aT18. In addition to free serine, five peptide fragments, S1, S2, S4, S5, and S6, were isolated by electrophoresis at pH 1.9. Their amino acid compositions are given in Table 11. The sequence of S4 was found to be *Ser-Val* by dansylation. S1 confirms the sequence of Th1 and S5 and S6 that of Th8. S4 gives the overlap between Th1 and Th4 as well as between Th6 and Th3. The overlap is further established by S2.

## DISCUSSION

The amino acid sequences of seventeen peptides isolated from the tryptic hydrolysates of the slow component of soybean leghemoglobin have been determined. With the exception of peptide *aT15*, all other peptides were obtained in sufficient amount to establish the sequence unequivocally. The proposed sequences depend on evidence obtained by enzymatic digestion of the tryptic peptides with chymotrypsin, subtilopectidase A, thermolysin, and by Edman sequential degradation. The presence of amides in aspartyl or glutamyl residues was determined from the behaviour of peptides on electrophoresis or by quantitative determination of free amino acids released by the action of leucine aminopeptidase or carboxypeptidase on the peptide containing the amino acids in question.

The nature of peptide bonds that are cleaved by trypsin conform to known specificity requirements of the enzyme. However, the rate of hydrolysis of various bonds differs. The presence of aspartic acid in a position  $\text{NH}_2$ -terminal to lysine (*-Asp-Lys-*) decreases the rate of hydrolysis. This sequence occurs in peptide *aT8*. The isolation of the ditryptic peptide, *aT8*, *aT16*, in good yield clearly indicates the resistance of this bond in contrast to peptides having glutamic acid in the same position (*-Glu-Lys-*), e.g. *aT11*, *aT12*, and *aT18*.

As was expected from the known specificity of chymotrypsin, cleavage occurred at peptide bonds involving the carboxyl groups of the tyrosine, phenylalanine, and leucine residues. The only exception was the tyrosyl-aspartyl bond in peptide *aT17*. In addition, it was observed that chymotrypsin hydrolyzed a particular peptide linkage at a much faster rate than other potentially sensitive ones. Hirs *et al.*<sup>20</sup> have suggested that a longer sequence may determine the ability of a particular bond to be cleaved. To what extent the conformation of the peptide might be decisive remained to be clarified.

Thermolysin was found to be highly specific in hydrolyzing bonds involving the amino groups of tyrosine, phenylalanine, leucine, valine, and isoleucine. However, an exceptional splitting at the carboxyl group of tyrosine was observed with one of the thermolytic subpeptides of *aT18*. A preference of some bonds was also observed with thermolysin. In peptide *aT16* the *-Gly-Val-* sequence was not attacked by the enzyme. In the case of thermolysin too it is evident that something more than the presence of a specific amino acid residue is required for the enzyme specificity.

The proposed amino acid sequence of the tryptic peptides is primarily based on the data of subtractive- and dansyl-Edman sequential techniques. The cyclization of the  $\text{NH}_2$ -terminal glutamine residue to pyrrolidone carboxylic acid, one of the sources of difficulty with the Edman procedure, was encountered only with peptide *aT13*. By separating the tryptic peptides electrophoretically, we were able to determine the  $\text{NH}_2$ -terminal glutamine residue before cyclization.

The peptide *aT15* is a minor component isolated from one leghemoglobin preparation only. Because the preparation of leghemoglobin requires the use of a pool of nodules from hundreds of plants the noted heterogeneity can be due to an intraspecies variation. Further sequence studies on the other soybean leghemoglobin components may shed some light on the origin of this leghemoglobin form.



*Acknowledgement.* This investigation has in part received financial support from the Finnish National Research Council for Sciences.

## REFERENCES

1. Ellfolk, N. and Sievers, G. *Acta Chem. Scand.* **26** (1972) 1155.
2. Ellfolk, N. *Acta Chem. Scand.* **14** (1960) 609.
3. Ellfolk, N. *Acta Chem. Scand.* **16** (1962) 831.
4. Ryle, A. P., Sanger, F., Smith, L. F. and Kitai, R. *Biochem. J.* **60** (1955) 541.
5. Ambler, R. P. *Biochem. J.* **89** (1963) 349.
6. Spackman, D. H., Stein, W. H. and Moore, S. *Anal. Chem.* **30** (1958) 1190.
7. Spies, J. R. and Chambers, D. C. *Anal. Chem.* **21** (1949) 1249.
8. Sjöquist, J., Blombäck, B. and Wallén, P. *Arkiv Kemi* **16** (1960) 425.
9. Blombäck, B., Blombäck, M., Edman, P. and Hessel, B. *Biochim. Biophys. Acta* **115** (1966) 371.
10. Jeppsson, J. D. and Sjöquist, J. *Anal. Biochem.* **18** (1967) 264.
11. Königsberg, W. and Hill, R. J. *J. Biol. Chem.* **237** (1962) 2547.
12. Königsberg, W. *J. Biol. Chem.* **241** (1966) 4534.
13. Gray, W. R. and Hartley, B. S. *Biochem. J.* **89** (1963) 59P.
14. Gray, W. R. and Hartley, B. S. *Biochem. J.* **89** (1963) 379.
15. Milstein, C. *Biochem. J.* **101** (1966) 338.
16. Woods, K. R. and Wang, K.-T. *Biochim. Biophys. Acta* **133** (1967) 369.
17. Offord, R. E. *Nature* **211** (1966) 591.
18. Ellfolk, N. and Sievers, G. *Acta Chem. Scand.* **23** (1969) 2994.
19. Aggarwal, S. J. and Riggs, A. *Acta Chem. Scand.* **24** (1970) 2234.
20. Hirs, C. H. W., Moore, S. and Stein, W. H. *J. Biol. Chem.* **235** (1960) 633.

Received March 30, 1973.