ACTA CHEMICA SCANDINAVICA 27 (1973) 3817–3831

The Primary Structure of Soybean Leghemoglobin

III. Fractionation and Sequence of Chymotryptic Peptides of the Apoprotein of the Slow Component

NILS ELLFOLK and GUNNEL SIEVERS

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

The peptides from the chymotryptic digest of the apoprotein of the slow component of soybean leghemoglobin were fractionated by chromatography on Dowex 1 and high voltage paper electrophoresis. A suitable combination of these methods enables the isolation of 23 peptides, the amino acid compositions of which have been determined. The amino acid sequence of the peptides forming bridges between tryptic peptides was determined. The present results allow the combination of the tryptic peptides into four fragments.

The isolation of seventeen tryptic peptides from the slow component (Lb*a*) of soybean leghemoglobin and their sequences have recently been reported.[1,2] A digestion of Lb*a* was performed with chymotrypsin in order to elucidate the complete primary structure of Lb*a* by combining the different tryptic peptides. This communication reports the fractionation, purification, and amino acid compositions of the chymotryptic peptides obtained. The determination of the amino acid sequence has been performed mainly for the bridge peptides. Non-bridge peptides provided valuable confirmatory data on the amino acid sequence of the tryptic peptides already worked out.

EXPERIMENTAL

Materials. The slow component (Lb*a*) of soybean leghemoglobin and its apoprotein were prepared as described previously.[3,4] Carboxypeptidase A was a crystallized, DFP-treated preparation, and subtilopeptidase 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.) Trypsin was a crystalline, chymotrypsin-free preparation from Serva AG (Heidelberg, West-Germany).

* *Abbreviation.* 1-Dimethyl-amino-5-naphthalenesulfonyl-, dansyl, DNS-.

Chymotryptic digestion. The apoprotein of Lb*a* (192 mg) was denatured by incubation at pH 9.7 for 4 min at 95°C. Digestion was performed in 20 ml of solution at pH 8.00 and 25°C. 3.55 mg of chymotrypsin in 1 ml of 0.001 N HCl were added initially. Hydrolysis was performed in a Radiometer pH-stat by continuous titration to pH 8.00 with 1 N NaOH. After hydrolysis (3 h) the pH of the solution was reduced to 5.8. The precipitate formed was removed by centrifugation and the supernatant was lyophilized.

Column chromatography of chymotryptic peptides. The lyophilized digest was dissolved in 5 ml of water, adjusted to pH 8.8 with 1 N NaOH and added to a column (1.5 × 80 cm) of Dowex 1 × 2 (200-400 mesh) equilibrated to pH 8.4 at 35°C in a buffer composed of 1 % 2,4,6-collidine and 1 % pyridine adjusted to pH 8.4 with glacial acetic acid. Elution was initiated with pH 8.4 buffer at a flow rate of 25 ml/h, and 2.5 ml fractions were collected. After fraction No. 24 a gradient was set up with 300 ml of pH 8.4 buffer in a closed mixing chamber and a buffer of 1 % 2,4,6-collidine and 1 % pyridine adjusted to pH 6.5 with glacial acetic acid. The pH 6.5 buffer was changed to 0.5 N acetic acid at fraction 144 and this to 1 N acetic acid at fraction 176 and, finally, at fraction No. 256 to glacial acetic acid. A 0.1 ml aliquot of each tube was hydrolyzed with alkali and analyzed by the ninhydrin method previously described.[1] The contents of the tubes representing each peak were combined, dried *in vacuo*, dissolved in 2 ml of water and examined for purity as described previously.[1]

Further purification of peptides. Heterogeneous peptide fractions were further fractionated into their components either by the paper chromatography or the high voltage paper electrophoresis that had revealed the heterogeneity. The paper chromatography was performed in the butanol – acetic acid – water system as previously described.[1] The electrophoresis was performed at buffers of pH 1.9, 3.5, 6.4, and 8.9, the compositions of which have been reported earlier.[1,2].

Amino acid composition of peptides. A measured aliquot of peptide solution was hydrolyzed in 2 ml of 6 N HCl in evacuated sealed tubes for 18 h and the amino acid

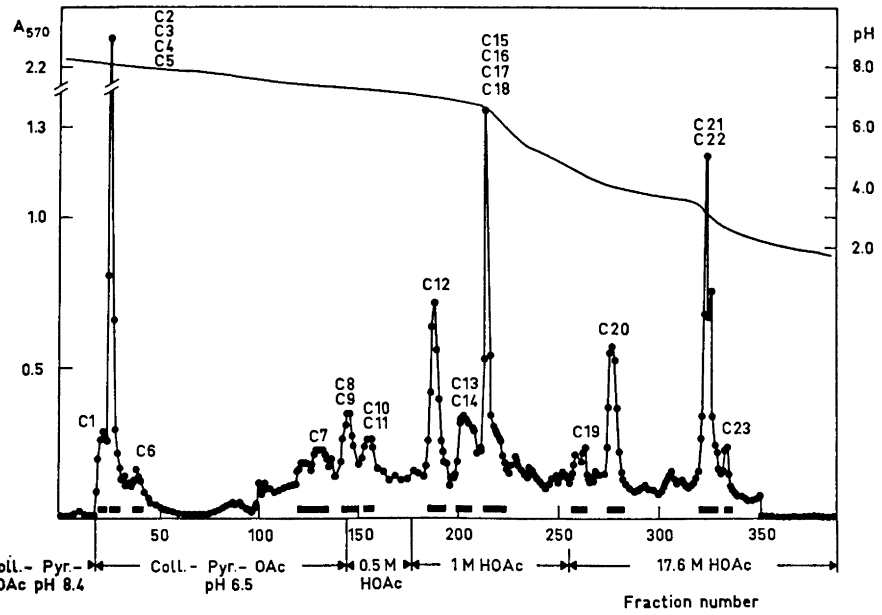

Fig. 1. Fractionation of peptides from a chymotryptic digest of 12.6 μmol of the slow component (Lb*a*) of soybean leghemoglobin on a Dowex 1 × 2 column (1.5 × 80 cm) at 35°C. The flow rate was 25 ml/h and 2.5 ml fractions were collected. Experimental details are given in the text. The pooled fractions are indicated by black bars.

Table 1. Amino acid compositions of the chymotryptic peptides of the slow component (Lb*a*) of soybean leghemoglobin. The values given in the table are mol ratios and are not corrected for losses of amino acids during hydrolysis.

Amino acid	*a*C1	*a*C2	*a*C3	*a*C4	*a*C5	*a*C6	*a*C7	*a*C8
Aspartic acid			1.03					1.76
Threonine			0.89					1.34
Serine							0.98	
Glutamic acid								1.98(3)
Proline								1.02
Glycine			0.95					
Alanine	3.82	2.82	2.03	2.98	0.80	1.07	1.60	3.49
Valine			0.95					2.02
Isoleucine	0.94	1.03	0.98	0.76	1.04			
Leucine								2.20
Tyrosine								
Phenylalanine								0.75
Tryptophan			+[a]				+[a]	
Lysine	2.24	1.15	3.19	2.27	2.15	0.93		2.17
Histidine								
Arginine							1.03	
No. of residues	7	5	11	6	4	2	4	17
E_{Lys}	0.69	0.50	0.50	0.76	0.89	0.78	0.50	neutr.
E_{Asp}								
R_{Leu}	0.10	0.40	0.11	0.41	0.29	0.19	0.45	0.22

[a] Positive Ehrlich reaction on paper.

Amino acid	*a*C9	*a*C10	*a*C11	*a*C12	*a*C13	*a*C14	*a*C15
Aspartic acid	1.15	1.06	1.03	1.03	1.13	0.91	0.91
Threonine	0.98		0.91			0.94	
Serine	2.01	0.89		1.00		0.90	
Glutamic acid		1.14	2.05	1.08	1.03	1.08	1.05
Proline			1.03		1.24	1.07	0.87
Glycine	2.06	1.01		1.06			
Alanine	3.93	2.01	2.14	1.18	2.82	3.03	1.11
Valine	2.16	0.99	2.89	0.91			
Isoleucine						0.96	0.93
Leucine	1.06	1.92	1.93	1.11	1.01	2.00	
Tyrosine							0.67
Phenylalanine			0.93		0.79	0.91	
Tryptophan							
Lysine	0.92		2.10		1.99	2.09	
Histidine	0.71						
Arginine		0.86		0.73			
No of residues	16	10	15	8	10	14	6
E_{Lys}	0.19	neutr.	neutr.	neutr.	neutr.	neutr.	neutr.
E_{Asp}							
R_{Leu}	0.38	0.62	0.68	0.37	0.24	0.12	0.39

Table 1. Continued.

Amino acid	*a*C16	*a*C17	*a*C18	*a*C19	*a*C20	*a*C21	*a*C22	*a*C23
Aspartic acid			2.95	1.05	1.05		1.06	1.12
Threonine			1.72	0.87	0.86			
Serine	2.97				2.42		1.78	
Glutamic acid			1.30	2.04	1.89	0.98	1.00	2.42
Proline			1.78					
Glycine			2.03					
Alanine		1.06	2.37	1.06	1.11	1.09	1.10	1.10
Valine	1.00	0.97	1.10		0.91			0.74
Isoleucine								
Leucine			2.83	0.97	0.95		0.98	0.88
Tyrosine								0.74
Phenylalanine	1.03	0.96	1.02		0.82	0.93		
Tryptophan							+[a]	
Lysine			1.95	1.02	1.00			
Histidine			0.75					
Arginine							0.87	
No. of residues	5	3	20	7	11	3	8	7
E_{Lys}								
E_{Asp}	neutr.	neutr.	neutr.	0.26	0.22	0.50	0.22	0.78
R_{Leu}	0.67	1.09	0.13	0.28	0.60	0.95	0.44	0.99

[a] Positive Ehrlich reaction on paper.

composition determined on the amino acid analyzer (Beckman/Spinco 120B). The Ehrlich reaction was employed to detect peptides containing tryptophan.[1] Amide residues were assigned on the basis of electrophoretic mobilities of the peptides at pH 6.5, according to Offord.[5]

Sequence studies. The chymotryptic peptides containing lysine or arginine residues were hydrolyzed with trypsin. A solution of the peptide, 1 μmol in 1 ml of 1 % ammonium bicarbonate buffer of pH 8.1 was treated with trypsin, 10 μl of a 1 % solution in 0.001 N HCl, and incubated at 37°C for 24 h. The resulting tryptic peptides were designated by the prefix T−. Digestion with subtilopeptidase A was carried out under the same experimental conditions as previously described.[2] The resulting peptides were designated by the prefix S–. Hydrolysis with leucine aminopeptidase and carboxypeptidase A was performed under the same conditions as reported earlier.[2] Subtractive- and dansyl-Edman degradation procedures were performed as described in the preceding paper.[2]

Nomenclature of peptides. The peptides isolated from the different fractions were numbered according to the same principles as described previously.[1] The prefix C is placed before peptide number to indicate that the peptide was obtained from a chymotryptic digest. "*a*" before C indicates Lb*a*, the slow component of soybean leghemoglobin.

RESULTS

Amino acid composition of chymotryptic peptides. Fig. 1 shows the elution pattern of the chymotryptic peptides from a Dowex 1 column. The fractions were combined, checked for homogeneity and, if necessary, further purified by high voltage paper electrophoresis. Homogeneous peptides were subjected to amino acid analysis yielding the results shown in Table 1.

Amino acid sequence of chymotryptic peptides. The deduced sequences of the chymotryptic peptides isolated are given in Table 2. Because the sequences of the tryptic peptides of Lb*a* have already been established,[2] those of most chymotryptic peptides could be evaluated from their amino acid compositions and their NH_2-terminal analyses. Structural studies were performed mainly on bridge-peptides containing lysine and arginine. In the succeeding section the peptides are considered individually and the evidence for their amino acid sequence is presented.

Table 2. Deduced amino acid sequences of the chymotryptic peptides from the slow component (Lb*a*) of soybean leghemoglobin.

Peptide number	Amino acid sequence
aC1	Ala-[Ala,Ala,Ile]-Lys-Ala-Lys
*a*C2	Ala-Ala-Ala-[Ile,Lys]
*a*C3	Lys-Thr-Ile-Lys-Ala-[Ala,Val,Gly,Asp]Lys-Trp
*a*C4	Ala-[Ala,Ile,Lys,Ala,Lys]
*a*C5	Ile-Lys-Ala-Lys
*a*C6	Ala-Lys
*a*C7	Ser-Arg-Ala-Trp
*a*C8	Ala-Gln-Lys-Ala-[Val,Thr,Asn,Pro,Glu,Phe,Val,Val]-Lys-Glu-Ala-Leu-Leu
*a*C9	Lys-Ala-[Ser,Gly,Thr,Val,Val,Ala,Asp,Ala,Ala,Leu,Gly,Ser,Val,His]
*a*C10	Ala-Leu-Val-Arg-Asp-Ser-[Ala,Gly,Gln]-Leu
*a*C11	Lys-Ala-Val-[Thr,Asn,Pro,Glu,Phe,Val,Val,Lys,Glu,Ala,Leu,Leu]
*a*C12	Val-Arg-Asp-Ser-Ala-[Gly,Gln]-Leu
*a*C13	Glu-Lys-Ala-[Pro,Ala,Ala]-Lys-Asp-[Leu,Phe]
*a*C14	Thr-Ser-Ile-Leu-Glu-Lys-Ala-Pro-Ala-Ala-Lys-Asp-Leu-Phe
*a*C15	Ala-[Asn,Ile,Pro,Gln,Tyr]
*a*C16	Val-Ser-Ser-Ser-Phe
*a*C17	Val-Ala-Phe
*a*C18	Leu-Ala-[Asn,Pro,Thr,Asp,Gly,Val,Asn,Pro]-Lys-Leu-[Thr,Gly,His,Ala, Glu]-Lys-Leu-Phe
*a*C19	Thr-[Glu,Lys,Gln,Asp,Ala]-Leu
*a*C20	Thr-Glu-Lys-Gln-[Asp,Ala,Leu,Val,Ser,Ser]-Ser-Phe
*a*C21	Glu-Ala-Phe
*a*C22	Ser-Asp-Glu-[Leu,Ser]-Arg-Ala-Trp
*a*C23	Glu-Val-[Ala,Tyr,Asp,Glu]-Leu

Peptide a*C1.* Alanine was established as the NH_2-terminal residue by dansylation. Tryptic digestion of *a*C1 produced two fragments, T1 and T2, which were isolated by electrophoresis at pH 6.5. The amino acid composition (Table 3) of T1 corresponds uniquely to the COOH-terminal portion of *a*T17 and that of T2 to *a*T4. Dansylation confirmed the sequence of T2 as *Ala-Lys.* Therefore, *a*C1 provides overlap joining *a*T17 to *a*T4.

Peptide a*C2.* Three steps of Edman degradation established the NH_2-terminal sequence *Ala-Ala-Ala-* (Table 3). *a*C2 represents the COOH-terminal portion of tryptic peptide *a*T17.

Peptide a*C3.* Three stages of Edman degradation indicated that the NH_2-terminal sequence is *Lys-Thr-Ile-*. Tryptic digestion of *a*C3 yielded four

Table 3. Amino acid sequences of peptides *a*C1, *a*C2, and *a*C3.

Sequence *a*C1	Ala-[Ala,Ala,Ile]-Lys-Ala-Lys
	|————T1————|-T2—|
Dansylation	DNS-Ala
Tryptic peptides	
T1	Ala,2.90; Ile,0.92; Lys,1.08
(E_{Lys},0.50)	
Dansylation	DNS-Ala
T2	Ala,1.03; Lys,0.97
(E_{Lys},0.79)	
Dansylation	DNS-Ala
Sequence *a*C2	Ala-Ala-Ala-[Ile,Lys]
Edman degradation	Direct examination of dansyl derivatives
Step 1	DNS-Ala
Step 2	DNS-Ala
Step 3	DNS-Ala
Sequence *a*C3	Lys-Thr-Ile-Lys-Ala-[Ala,Val,Gly,Asp]-Lys-Trp
	|-T4-|—T3—|————T2————|-T1-|
Edman degradation	Direct examination of dansyl derivatives
Step 1	DNS-Lys
Step 2	DNS-Thr
Step 3	DNS-Ile
Tryptic peptides	
T1	Trp +[a])
Dansylation	DNS-Trp
T2	Ala,1.77; Val,1.02; Gly,1.00; Asp,1.15; Lys,1.03
(neutral)	
Dansylation	DNS-Ala
T3	Thr,0.73; Ile,1.00; Lys,1.03
(E_{Lys},0.58)	
Dansylation	DNS-Thr
T4	Lys,1.00
(E_{Lys},1.00)	

[a] Positive Ehrlich reaction on paper.

fragments, T1, T2, T3, and T4, which were purified by electrophoresis at pH 6.5 and 1.9 (Table 3). T1 represented free tryptophan as identified by paper electrophoresis and paper chromatography. The amino acid compositions of T2 and T3 were compatible with those of *a*T8 and *a*T5, respectively. T4 was found to be free lysine. Three tryptic peptides have COOH-terminal sequences, which should yield free lysine on chymotryptic hydrolysis, *a*T9, *a*T10, and *a*T13. Because tryptophan is the unique NH_2-terminal amino acid residue of *a*T16, *a*C3 provides the overlap establishing the tryptic peptide order *a*T5-*a*T8-*a*T16.

Peptide a*C4 and* a*C5.* Both these peptides confirm the tryptic peptide linkage *a*T17 to *a*T4. Three steps of Edman degradation established the sequence *Ile-Lys-Ala-Lys.*

Peptide a*C6.* One step of Edman degradation established the sequence *Ala-Lys.*

Peptide a*C7*. Dansylation established the NH_2-terminal serine residue. A COOH-terminal tryptophan residue is consistent with the specificity of chymotrypsin. Tryptic digestion of *a*C7 yielded two fragments, T1 and T2, which were isolated by electrophoresis at pH 6.5. Their amino acid composition is given in Table 4. The sequence of T1 was found to be *Ala-Trp* and that of T2 *Ser-Arg*. Peptide *a*C7, therefore, represents the overlap between *a*T16 and *a*T17.

Table 4. Amino acid sequences of peptides *a*C4, *a*C5, *a*C6, and *a*C7.

Sequence *a*C4	Ala-[Ala,Ile,Lys,Ala,Lys]
Dansylation	DNS-Ala
Sequence *a*C5	Ile-Lys-Ala-Lys
Edman degradation	Direct examination of dansyl derivatives
Step 1	DNS-Ile
Step 2	DNS-Lys
Step 3	DNS-Ala
Sequence *a*C6	Ala-Lys
Edman degradation	
Step 1	Ala, 0; Lys, 1.0
Sequence *a*C7	Ser-Arg-Ala-Trp |—T2—|—T1—|
Dansylation	DNS-Ser
Tryptic peptides	
T1	Ala, 1.00; Trp, +[a]
(neutral)	
Dansylation	DNS-Ala
T2	Ser, 1.02; Arg, 0.98
(E_{Lys}, 0.78)	
Dansylation	DNS-Ser

[a] Positive Ehrlich reaction on paper.

Peptide a*C8*. Alanine was the NH_2-terminal amino acid residue as found by dansylation. Digestion of *a*C8 with trypsin produced three fragments, T1, T2, and T3, which were isolated by electrophoresis at pH 6.5 and their amino acid composition is given in Table 5. Two steps of Edman degradation showed T1 to have the sequence *Glu-Ala-Leu-Leu* corresponding to residues 1 – 4 of the tryptic peptide *a*T10. The amino acid composition and analysis of the NH_2-terminal amino acid shows T2 to be identical with *a*T7. The sequence of T3 was *Ala-Gln-Lys*, which is compatible with the COOH-terminal sequence of *a*T6. Thus, peptide *a*C8 unambiguosly indicates the tryptic peptide order *a*T6-*a*T7-*a*T10.

Peptide a*C9*. Dansylation showed lysine as the NH_2-terminal amino acid residue. Tryptic digestion produced two fragments, T1 and T2, which were separated on electrophoresis at pH 6.5. Their amino acid contents are given in Table 5. T1 represents the NH_2-terminal portion of *a*T6 as decided from its

amino acid composition and determination of the NH_2-terminal amino acid residue. T2 was found to be free lysine. Three tryptic peptides, *a*T9, *a*T10, and *a*T13, have COOH-terminal sequences which allow liberation of free lysine on hydrolysis with chymotrypsin. Of these, *a*T10 is excluded because of its position COOH-terminal to *a*T7.

Table 5. Amino acid sequences of peptides *a*C8 and *a*C9.

Sequence *a*C8	Ala-Gln-Lys-Ala-[Val,Thr,Asn,Pro,Glu,Phe,Val,Val]-Lys-
	|—T3—|————————T2————|
	Glu-Ala-Leu-Leu
	|——T1———|
Dansylation	DNS-Ala
Tryptic peptides	
T1	Glu, 1.07; Ala, 1.17; Leu, 1.76
(E_{Asp}, 0.48)	
Edman degradation	
Step 1	*Glu, 0.57*; Ala, 1.18; Leu, 1.82
Step 2	*Ala, 0.68*; Leu, 2.00
T2	Ala, 1.08; Val, 2.56; Thr, 1.00; Asp, 1.17; Pro, 0.70; Glu, 1.11
(neutral)	Phe, 0.71; Lys, 0.98
Dansylation	DNS-Ala
T3	Ala, 1.10; Glu, 1.00; Lys, 0.90
(E_{Lys}, 0.58)	
Dansylation	DNS-Ala
Sequence *a*C9	Lys-Ala-[Ser,Gly,Thr,Val,Val,Ala,Asp,Ala,Ala,Leu,Gly,
	|-T2-|————————T1————
	Ser,Val,His]
	————|
Dansylation	DNS-Lys
Tryptic peptides	
T1	Ala, 3.81; Ser, 2.25; Gly, 1.94; Thr, 0.75; Val, 2.00(3);
(E_{Asp},0.24)	Asp, 1.10; Leu, 1.09; His, 0.95
Dansylation	DNS-Ala
T2	Lys, 1.00
(E_{Lys}, 1.00)	
Dansylation	DNS-Lys

Peptide a*C10.* Alanine was established as the NH_2-terminal amino acid residue. A COOH-terminal leucine residue was assigned on the basis of the specificity of chymotrypsin. Two fragments, T1 and T2, were obtained on hydrolysis with trypsin and isolated by electrophoresis at pH 6.5. Their amino acid contents are given in Table 6. Two steps of Edman degradation on T1 showed the sequence *Asp-Ser-(Ala, Gly, Gln)-Leu*, which is compatible with the NH_2-terminal sequence of *a*T9. Two steps of Edman degradation on T2 established the sequence *Ala-Leu-Val-Arg*, representing the COOH-terminal sequence of *a*T3. *a*C10, therefore, provides the overlap joining *a*T3 to *a*T9.

Peptide a*C11.* This peptide (Table 6) provided further evidence of the linkage between the tryptic peptides *a*T7 and *a*T10.

Peptide a*C12.* Dansylation revealed valine as the NH_2-terminal amino acid residue. A COOH-terminal leucine residue was assigned on the basis of

Table 6. Amino acid sequences of peptides *a*C10 and *a*C11.

Sequence *a*C10	Ala-Leu-Val-Arg-Asp-Ser-[Ala,Gly,Gln]Leu |———T2———|————————T1————|
Dansylation	DNS-Ala
Tryptic peptides	
T1	Asp, 0.99; Ser, 0.86; Ala, 1.02; Gly, 1.01; Glu, 1.11; Leu, 1.03
(E_{Asp},0.43)	
Edman degradation	Direct examination of dansyl derivatives
Step 1	DNS-Asp
Step 2	DNS-Ser
T2	Ala, 1.16; Leu, 0.99; Val, 1.01; Arg, 0.90
(E_{Lys},0.52)	
Edman degradation	Direct examination of dansyl derivatives
Step 1	DNS-Ala
Step 2	DNS-Leu
Sequence *a*C11	Lys-Ala-Val-[Thr,Asn,Pro,Glu,Phe,Val,Val,Lys,Glu,Ala, Leu,Leu]
Edman degradation	Direct examination of dansyl derivatives
Step 1	DNS-Lys
Step 2	DNS-Ala
Step 3	DNS-Val

Table 7. Amino acid sequences of peptides *a*C12 and *a*C13.

Sequence *a*C12	Val-Arg-Asp-Ser-Ala-[Gly,Gln,Leu] |—T2—|————————T1————|
Dansylation	DNS-Val
Tryptic peptides	
T1	Asp, 0.84; Ser, 0.87; Ala, 1.03; Gly, 0.97; Glu, 1.01; Leu, 1.12
(E_{Asp},0.43)	
Edman degradation	Direct examination of dansyl derivatives
Step 1	DNS-Asp
Step 2	DNS-Ser
Step 3	DNS-Ala
T2	Val, 1.03; Arg, 0.97
(E_{Lys},0.80)	
Dansylation	DNS-Val
Sequence *a*C13	Glu-Lys-Ala-[Pro,Ala,Ala]-Lys-Asp-[Leu,Phe] |————————T2————————|——T1——| |———T3————————|
Dansylation	DNS-Glu
Tryptic peptides	
T1	Asp, 1.12; Leu, 1.05; Phe, 0.82
(E_{Asp},0.51)	
Dansylation	DNS-Asp
T2	Glu, 0.96; Lys, 1.97; Ala, 3.04; Pro, 1.02
(E_{Lys},0.38)	
Dansylation	DNS-Glu
T3	Ala, 2.95; Pro, 1.07; Lys, 0.98
(E_{Lys},0.55)	
Dansylation	DNS-Ala

the specificity of chymotrypsin. Hydrolysis with trypsin yielded two fragments, T1 and T2, which were isolated by electrophoresis at pH 6.5 and had the amino acid contents given in Table 7. Peptide *a*C12 provided further support for the linkage of the tryptic peptides *a*T3 and *a*T9.

Peptide a*C13.* Glutamic acid was the NH_2-terminal amino acid residue as found by dansylation. Tryptic hydrolysis produced three fragments, T1, T2, and T3, which were isolated by electrophoresis at pH 6.5 and had the amino acid compositions given in Table 7. T1 represents the unique NH-$_2$terminal portion of the tryptic peptide *a*T14, and T3 that of the tryptic peptide *a*T2. This indicates the tryptic peptide order *a*T18-*a*T2-*a*T14, which was further confirmed by the chymotryptic peptide *a*C14.

Peptide a*C14.* Edman degradation and leucine amino peptidase digestion gave the NH_2-terminal sequence *Thr-Ser-Ile-Leu-Glu-* (Table 8). Carboxypeptidase A digestion revealed the COOH-terminal sequence *-Leu-Phe.* Tryptic digestion yielded three fragments, T1, T2, and T3, which were isolated on electrophoresis at pH 6.5 and 1.9 and found to have the amino acid compositions given in Table 8. By Edman degradation the sequence of T1 was found to be *Asp-Leu-Phe-*, the unique NH_2-terminal sequence of tryptic peptide *a*T14, and that of T2 *Thr-Ser-Ile-Leu-Glu-Lys* the COOH-terminal

Table 8 Amino acid sequences of peptides *a*C14 and *a*C15.

Sequence *a*C14	Thr-Ser-Ile-Leu-Glu-Lys-Ala-Pro-Ala-Ala-Lys-Asp-Leu-Phe
	|————T2————|————T3————|———T1——|
Edman degradation	Direct examination of dansyl derivatives
Step 1	DNS-Thr
Step 2	DNS-Ser
Step 3	DNS-Ile
Leucine amino-peptidase, 30 min	Thr, 1.00; Ser, 1.00; Ile, 0.48; Leu, 0.45; Glu, 0.28
Carboxypeptidase A 20 min.	Leu, 0.60; Phe, 1.00
Tryptic peptides	
T1 (E_{Asp},0.51)	Asp, 1.04; Leu, 1.02; Phe, 0.95
Edman degradation	Direct analysis of dansyl derivatives
Step 1	DNS-Asp
Step 2	DNS-Leu
T2 (neutral)	Thr, 0.92; Ser, 1.10; Ile, 0.90; Leu, 0.91; Glu, 1.02; Lys, 1.16
Edman degradation	Direct examination of dansyl derivatives
Step 1	DNS-Thr
Step 2	DNS-Ser
Step 3	DNS-Ile
T3 (E_{Lys},0.52)	Ala, 2.95; Pro, 1.07; Lys, 0.98
Edman degradation	Direct examination of dansyl derivatives
Step 1	DNS-Ala
Step 2	DNS-Pro
Sequence *a*C15	Ala-[Asn,Ile,Pro,Gln,Tyr]
Dansylation	DNS-Ala

portion of the tryptic peptide *a*T18. T3 was found to be the tryptic peptide *a*T2. These results indicate the tryptic peptide order *a*T18-*a*T2-*a*T14.

Peptide a*C15*. Dansylation showed alanine to be the NH_2-terminal residue (Table 8). Peptide *a*C15 represents the NH_2-terminal sequence of the tryptic peptide *a*T18.

Peptide a*C16*. Three steps of Edman degradation revealed the NH_2-terminal sequence *Val-Ser-Ser-* (Table 9). A COOH-terminal phenylalanine residue was assigned on the basis of chymotryptic specificity. The sequence of *a*C16 was concluded to be *Val-Ser-Ser-Ser-Phe*, representing the middle portion of the tryptic peptide *a*T13.

Peptide a*C17*. Two steps of Edman degradation established the sequence *Val-Ala-Phe* (Table 9), which is the NH_2-terminal portion of the tryptic peptide *a*T12 and the Lb*a* chain.

Table 9. Amino acid sequences of peptides *a*C16, *a*C17, and *a*C18.

Sequence *a*C16	Val-Ser-Ser-[Ser,Phe]
Edman degradation	Direct examination of dansyl derivatives
Step 1	DNS-Val
Step 2	DNS-Ser
Step 3	DNS-Ser
Sequence *a*C17	Val-Ala-Phe
Edman degradation	Direct examination of dansyl derivatives
Step 1	DNS-Val
Step 2	DNS-Ala
Sequence *a*C18	Leu-Ala-[Asn,Pro,Thr,Asp,Gly,Val,Asn,Pro]-Lys- |————————T1————————|
	Leu-[Thr, Gly, His, Ala, Glu]-Lys-Leu-Phe |————————T3————————|——T2—|
Dansylation	DNS-Leu
Tryptic peptides	
T1	Leu, 0.78; Ala, 1.07; Asp, 2.98; Pro, 2.04; Thr, 1.00; Gly, 1.11;
(E_{Asp},0.28)	Val, 0.97; Lys, 1.03
Edman degradation	Direct examination of dansyl derivatives
Step 1	DNS-Leu
Step 2	DNS-Ala
T2	Leu, 1.06; Phe, 0.94
(neutral)	
Edman degradation	Direct examination of dansyl derivatives
Step 1	DNS-Leu
Step 2	DNS-Phe
T3	Leu, 0.96; Thr, 0.97; Gly, 1.04; His, 0.93; Ala, 1.03; Glu, 1.02; Lys, 0.99
Dansylation	DNS-Leu

Peptide a*C18*. Dansylation showed leucine as the NH_2-terminal residue. Hydrolysis with trypsin yielded three fragments, T1, T2, and T3, which were electrophoretically isolated at pH 6.5 and 1.9. The amino acid content of these fractions is given in Table 9. Two steps of Edman degradation on T1 indicated the sequence *Leu-Ala-* which, in addition to the amino acid content,

shows T1 to represent the COOH-terminal portion of the tryptic peptide *a*T14. The sequence of T2 was found to be *Leu-Phe* by dansylation, and represents the unique NH_2-terminal sequence of the tryptic peptide *a*T3. The amino acid composition of T3 was compatible with that of the tryptic peptide *a*T11. Therefore, peptide *a*C18 provides evidence of the tryptic peptide order *a*T14-*a*T11-*a*T3.

Peptide a*C19*. Threonine was established as the NH_2-terminal residue and leucine as the COOH-terminal residue using carboxypeptidase A (Table 10). *a*C19 indicates the tryptic peptide sequence *a*T12–*a*T13, which was further confirmed by peptide *a*T20.

Peptide a*C20*. Leucine amino peptidase digestion established the NH_2-terminal sequence as *Thr-Glu-*. The phenylalanine residue was positioned at the COOH terminus from chymotryptic specificity, and the action of carboxy-

Table 10. Amino acid sequences of peptides *a*C19, *a*C20, and *a*C21.

Sequence *a*C19	Thr-[Glu,Lys,Gln,Asp,Ala]-Leu
Dansylation	DNS-Thr
Carboxypeptidase A 20 min	Leu, 1.0
Sequence *a*C20	Thr-Glu-Lys-Gln-[Asp,Ala,Leu,Val,Ser,Ser]-Ser-Phe |——T2——|————————T1————|
Dansylation	DNS-Thr
Leucine amino-peptidase, 30 min	Thr, 1.00; Glu, 0.48
Carboxypeptidase A 30 min	Phe, 1.00; Ser, 0.47
Tryptic peptides	
T1 (E_{Asp},0.31)	Glu, 1.03; Asp, 0.95; Ala, 1.33; Leu, 1.07; Val, 1.08; Leu, 1.07; Phe, 0.86
Dansylation	DNS-Glu
T2 (neutral)	Thr, 0.91; Glu, 1.04; Lys, 1.05
Dansylation	DNS-Thr
Subtilopeptidase A peptides	
S1 (E_{Asp},0.55)	Asp, 0.98; Ala, 1.03; Leu, 0.98
Edman degradation	
Step 1	*Asp, 0.00*; Ala,0.98; Leu, 1.02
Step 2	*Ala, 0.00* Leu, 1.00
S2 (neutral)	Thr, 0.94; Glu, 2.06; Lys, 1.00
S3 (neutral)	Val, 1.14; Ser, 2.80
Edman degradation	
Step 1	*Val, 0.00* Ser, 3.00
Sequence *a*C21	Glu-Ala-Phe
Edman degradation	Direct examination of dansyl derivatives
Step 1	DNS-Glu
Step 2	DNS-Ala
Step 3	DNS-Phe

peptidase A on *a*C20 confirmed the COOH-terminal sequence as *-Ser-Phe*. *a*C20 was redigested with trypsin and two fragments were obtained, T1 and T2, which were isolated on electrophoresis at pH 6.5. The amino acid composition of T1 and T2 is given in Table 10. Dansylation on T2 established the sequence *Thr-Glu-Lys* for the neutral peptide. To confirm the sequence and the position of the amide, *a*C20 was subjected to hydrolysis with subtilopeptidase A, whereupon three major fragments, S1, S2, and S3, were isolated by electrophoresis at pH 6.5 and 1.9. Two steps of Edman degradation on S1 gave the sequence *Asp-Ala-Leu*. S2 was found to be neutral at pH 6.5 indicating that the second glutamic acid was amidated, and the sequence was concluded to be *Thr-Glu-Lys-Gln*. One step of Edman degradation for S3 gave the sequence *Val-Ser-Ser-Ser*. Thus the tryptic peptide sequence of *a*T12–*a*T13 is indicated.

Peptide a*C21*. Three steps of Edman degradation established the sequence *Glu-Ala-Phe* (Table 10), which represents residues 10 – 12 of the COOH-terminal portion of the tryptic peptide *a*T13.

Peptide a*C22*. Three steps of Edman degradation established the NH_2-terminal sequence as *Ser-Asp-Glu-*. A COOH-terminal tryptophan was assigned on the basis of the specificity of chymotrypsin. Hydrolysis with trypsin yielded two fragments, T1 and T2, which were isolated on electrophoresis at pH 6.5 and 1.9. Their amino acid contents are given in Table 11.

Table 11. Amino acid sequences of peptides *a*C22 and *a*C23.

Sequence *a*C22	Ser-Asp-Glu-[Leu,Ser]-Arg-Ala-Trp |——————T1——————|—T2—|
Edman degradation	
Step 1	*Ser, 1.08* Asp, 1.00; Glu, 0.97; Leu, 0.98; Arg, +; Ala, 0.97
Step 2	Ser, 1.05; *Asp, 0.25* Glu, 0.95; Leu, 1.02; Arg, +; Ala, 0.98
Step 3	Ser, 1.09; Asp, 0.00; *Glu, 0.39* Leu, 0.89; Arg +; Ala, 1.03
Tryptic peptides	
T1 (E_{Asp},0.34)	Ser, 1.79; Asp, 1.05; Glu, 1.02; Leu, 1.07; Arg, 1.06
Edman degradation	Direct examination of dansyl derivatives
Step 1	DNS-Ser
Step 2	DNS-Asp
T2 (neutral)	Ala, 1.00; Trp, +[a])
Edman degradation	Direct examination of dansyl derivatives
Step 1	DNS-Ala
Step 2	DNS-Trp
Sequence *a*C23	Glu-Val-[Ala,Tyr,Asp,Glu]-Leu
Edman degradation	
Step 1	DNS-Glu
Step 2	DNS-Val
Carboxypeptidase A	
30 min	Leu, 1.0

[a] Positive Ehrlich reaction on paper.

Two steps of Edman degradation showed the sequence of T2 to be *Ala-Trp*. Peptide *a*C22 provides further support for the linkage of *a*T16 to *a*T17.

Peptide a*C23*. This peptide is assumed to represent the NH_2-terminal portion of the tryptic peptide *a*T17, corresponding to residues 2 – 8 on the basis of the amino acid composition and NH_2-terminal analysis (Table 11).

DISCUSSION

The amino acid sequence of all the peptides obtained from the chymotryptic digest of Lb*a* could be evaluated from the direct evidence obtained in the present study and the known sequences of tryptic peptides.

With exception of some peptides ending in lysine, the COOH-terminal residues of the chymotryptic peptides studied seemed to be compatible with the known specificity of chymotrypsin. Peptides *a*C1, *a*C3, and *a*C4 all end in *Ala-Lys*, thus indicating a linkage between *a*T17 and *a*T4. Therefore *a*T17-*a*T4 is assumed to represent the COOH-terminus of Lb*a*. Peptide *a*C2 represents the COOH-terminus of *a*T17 with isoleucine next to lysine. Peptide *a*C15 also indicates a hydrolysis at the Lys-Ala bond, in which lysine is evidently preceded by a phenylalanine residue. Chymotryptic hydrolysis of a bond involving the carboxyl group of lysine is unusual, but it has been observed in a Lys-Ala bond in the γ-chain of human hemoglobin,[6] where the sequence *Val-Lys-Ala-His* was hydrolyzed.

The present study on the chymotryptic peptides permits the arrangement of the tryptic peptides into four fragments: *a*T12-*a*T13, *a*T18-*a*T2-*a*T14-*a*T11-*a*T3-*a*T9, *a*T6-*a*T7-*a*T10, and *a*T5-*a*T8-*a*T16-*a*T17-*a*T4. *a*T12-*a*T13 and *a*T5-*a*T8-*a*T16-*a*T17-*a*T4 represent, respectively, the NH_2- and COOH-terminal portions of Lb*a*. The two additional fragments both represent the middle portion of the peptide chain of Lb*a*. However, their internal order cannot be decided simply on the basis of the chymotryptic peptides isolated. If *a*T13 is assumed to be linked to *a*T18, then the distance between the two hemin binding histidines includes 30 amino acid residues, and if *a*T13 were bound to *a*T15 this separation would be 56 amino acid residues. In vertebrate hemoglobin chains the distance is 28 residues. Therefore *a*T13 is assumed to be linked to *a*T18. The lack of direct evidence for the attachment of *a*T13 to *a*T18, *a*T9 to *a*T6, and *a*T10 to *a*T5 is a result of the liberation of lysine as the NH_2-terminal residue on chymotryptic attack on these peptides because of their COOH-terminal portions *-Phe-Lys* for *a*T13, and *Leu-Lys* for *a*T9 and *a*T10, respectively. Complementary information is provided by isolating the overlapping peptides from a thermolytic digest of Lb*a*, the description of which is given in the next paper of this series.

The position of many of the chymotryptic peptides that did not contain lysine or arginine was deduced on the basis of their amino acid composition. Thus we were able to account for all the residues in the Lb*a* chain in terms of pure chymotryptic peptides, with the exception of a portion of *a*T18 and *a*T14. These results have led to the tentative formulation of the Lb*a* chain sequence shown in Fig. 2.

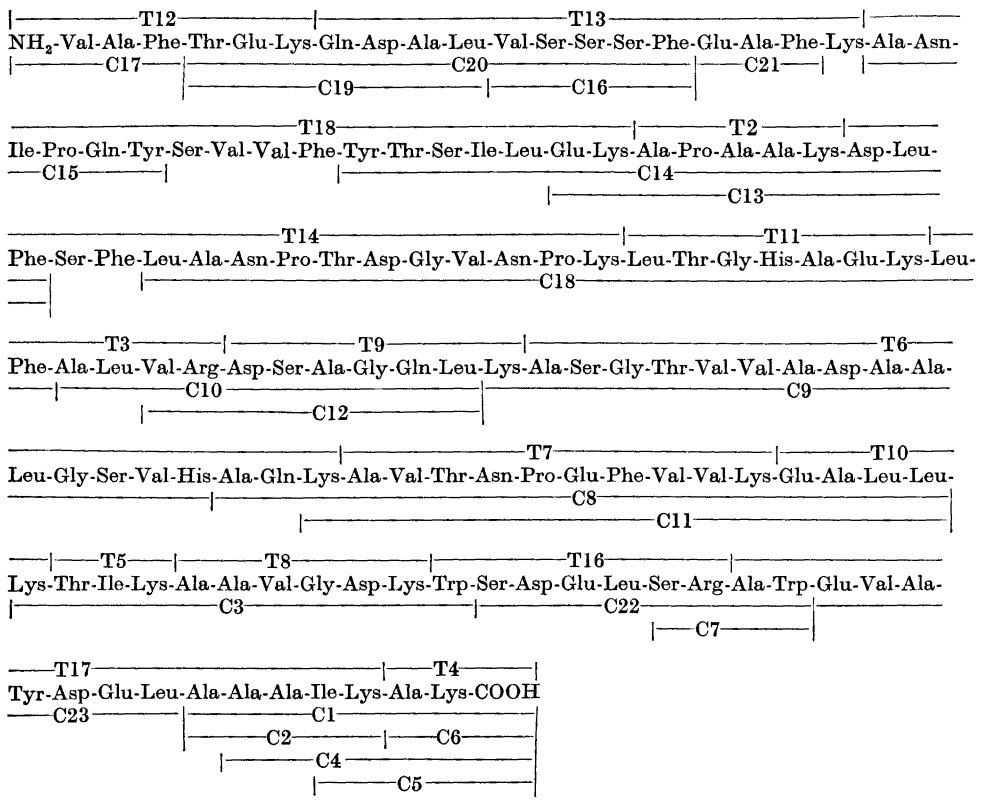

Fig. 2. The proposed amino acid sequence of soybean leghemoglobin *a*. The tryptic peptides (T) are shown above the sequence and the chymotryptic peptides (C) are shown below it.

Acknowledgement. This investigation has in part recieved 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. and Sievers, G. *Acta Chem. Scand.* **27** (1973) 3371.
3. Ellfolk, N. *Acta Chem. Scand.* **14** (1960) 609.
4. Ellfolk, N. *Acta Chem. Scand.* **16** (1962) 831.
5. Offord, R. E. *Nature* **211** (1966) 591.
6. Schroeder, W. A., Shelton, J. R., Shelton, J. B., Cormick, J. and Jones, R. T. *Biochemistry* **2** (1963) 992.

Received May 31, 1973.