ACTA CHEMICA SCANDINAVICA 27 (1973) 3986–3992

The Primary Structure of Soybean Leghemoglobin

IV. Fractionation and Sequence of Thermolytic Peptides of the Apoprotein of the Slow Component (Lb*a*)

NILS ELLFOLK and GUNNEL SIEVERS

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

The peptides from the thermolytic digest of the apoprotein of soybean leghemoglobin were fractionated by high voltage paper electrophoresis. Partial sequences of some of these peptides were obtained. Most of the overlap peptides were isolated and it was thus possible to order the tryptic peptides and to propose a unique sequence of the Lb*a* chain.

In the preceding paper[1] we reported that sixteen tryptic peptides[2,3] of the slow component of soybean leghemoglobin (Lb*a*) could be combined into four fragments by studying the chymotryptic peptides. Two of these fragments could be assigned to the NH_2- and the COOH-terminal portions of the entire molecule of Lb*a*. No direct evidence on the internal order of the other two fragments was obtained. Their order was decided by the distance between the two histidine residues, which was either 30 or 59 amino acid residues depending on the arrangement of the two fragments. A distance of 28 – 30 amino acid residues between the two hemin binding histidine residues is a general feature of hemoglobins and myoglobins.[4]

Because the specificity of thermolysin is known to differ from that of chymotrypsin, in the present experiments the thermolytic peptides of Lb*a* were isolated in order to obtain direct evidence for the internal arrangement of the two middle fragments mentioned above. Further, it was hoped that the peptides resulting from the thermolytic digestion of Lb*a* would enable us to derive the order of the tryptic peptides unequivocally and independently of the results of the chymotryptic digestion.

MATERIALS AND METHODS

Materials. The apoprotein of the slow component of soybean leghemoglobin (Lb*a*) was prepared as described previously.[5,6] Thermolysin was a crystalline preparation from Merck AG (West Germany). Trypsin was a crystalline chymotrypsin-free preparation

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

obtained from Serva AG (Heidelberg, West Germany). Carboxypeptidase A was a crystallized, DFP-treated preparation from Sigma Chemical Company (St Louis, U.S.A.). All other materials were the same as described previously.[2,3]

Thermolytic digestion. Digestion of the heat denatured (95°C, 3 min) apoLb*a* was carried out at 37°C in a pH-stat (Radiometer TT1) at pH 8.0. ApoLb*a* was suspended in water at a concentration of 1 %, and dissolved by the addition of 1 N NaOH. Thermolysin (0.5 mg/ml in 0.001 M calcium acetate) was added to give a final concentration of 0.5 % (w/w) and digestion was allowed to proceed for 2 h. A second aliquot of enzyme was then added and digestion was continued for a further hour. The reaction was stopped by lowering the pH to 2.25 by addition of 6 N HCl. The digest was lyophilized.

Isolation of thermolytic peptides. The lyophilized digest was dissolved in 5 ml of water and the pH adjusted to 5 with 1 N NaOH. In contrast to the isolation of the tryptic and chymotryptic peptides, the thermolytic peptides were purified on high voltage paper electrophoresis at pH 6.5, 3.5, 1.9, and 8.9 as described previously.[2,3] A butanol – acetic acid – water (4:1:5) system was used for the paper chromatography.

Amino acid analysis. Quantitative and qualitative amino acid analyses were performed as described earlier.[2]

Determination of NH_2-terminus and sequence. The NH_2-terminal residues of the peptides were determined by the dansyl procedure and the derivatives identified by electrophoresis and chromatography, as described in the preceding paper.[1] Dansyl- and subtractive-Edman procedures, performed as described before,[3] were used to obtain the sequence. Some of the thermolytic peptides containing lysine were hydrolyzed with trypsin as described earlier.[1] Hydrolysis with carboxypeptidase A was performed as described previously.[3]

Amide residues. Amide residues were assigned according to Offord on the basis of electrophoretic mobilities of the peptides at pH 6.5.[7]

Nomenclature of thermolytic peptides. The principles employed for numbering the peptides were identical with those used previously.[2] The thermolytic peptides are represented by the letters Th followed by a number indicating their order from the anode on electrophoresis. *a* before Th indicates Lb*a*.

RESULTS AND DISCUSSION

The peptides from the thermolytic digestion of Lb*a* were separated by repeated high voltage paper electrophoresis at different pH-values. Peptides found homogeneous on electrophoresis and chromatography in the butanol-acetic acid-water system were subjected to quantitative amino acid analysis with the results given in Table 1. Partial sequences were determined for some of these peptides by Edman degradation. These results are summarized in Table 2. For the sake of brevity, the data obtained after the application of only one or two steps of dansyl-Edman degradation have been omitted.

It was recently found [1] that the tryptic peptides could be arranged into four fragments by the overlap peptides from a chymotryptic digestion. The sequences of the thermolytic overlap peptides, combining these fragments in an unequivocal fashion, is given below.

Peptide a*Th21.* Dansylation showed phenylalanine to be the NH_2-terminal residue (Table 3). Two fragments, T1 and T2, were obtained on hydrolysis with trypsin. They were isolated by electrophoresis at pH 6.5 and their amino acid content is given in Table 3. Five steps of Edman degradation and digestion with carboxypeptidase A showed T1 to have the sequence *Ala-Asn-Ile-Pro-Gln-Tyr-Ser*, which is compatible with the NH_2-terminal sequence of the tryptic peptide *a*T18. NH_2-terminal analysis on T2 established the sequence *Phe-Lys*, representing the unique COOH-terminal sequence of the tryptic peptide *a*T13. *a*Th21, therefore provides, the overlap joining *a*T13 to *a*T18.

Table 1. Amino acid compositions of the thermolytic peptides of the slow component not corrected for losses of amino

Amino acid	*a*Th1	*a*Th2	*a*Th3	*a*Th4	*a*Th5	*a*Th6	*a*Th7
Aspartic acid	1.05	1.10		2.11	1.15	1.05	2.06
Threonine					0.80	0.94	
Serine				0.90			1.81
Glutamic acid	1.07	1.17	1.08	1.09	2.03	2.23	1.41
Proline							
Glycine				1.00			0.94
Alanine	1.01	1.90	1.00	0.92	1.08	1.00	1.81
Valine	1.00	0.96		0.98			0.91
Isoleucine							
Leucine		0.87					1.05
Tyrosine	0.87	0.32					
Phenylalanine			0.92			0.89	
Tryptophan				+[a]			+[a]
Lysine				1.01	0.94	0.89	1.17
Histidine							
Arginine							0.79
No. of residues	5	7	3	9	6	7	13
E_{Lys} / E_{Asp}	0.63	0.57	0.50	0.42	0.28	0.28	0.28

[a] Positive Ehrlich reaction on paper.

Amino acid	*a*Th14	*a*Th15	*a*Th16	*a*Th17	*a*Th18	*a*Th19
Aspartic acid	1.15					
Threonine		0.87				
Serine	0.96	0.98				
Glutamic acid	1.05	1.11		0.93		
Proline		0.93				
Glycine	1.03					
Alanine	1.03	1.16	2.96	0.85	1.07	1.03
Valine	0.97			1.44	0.93	
Isoleucine		0.89				
Leucine	0.88	0.99	1.01	1.03		1.90
Tyrosine						
Phenylalanine						1.00
Tryptophan						
Lysine		1.05		1.00		
Histidine						
Arginine	0.93					
No. of residues	8	8	4	6	2	4
E_{Lys} / E_{Asp}	neutr.	neutr.	neutr.	neutr.	neutr.	neutr.

(Lb*a*) of soybean leghemoglobin. The values given in the table are mol ratios and are acids during hydrolysis.

*a*Th8	*a*Th9	*a*Th10	*a*Th11	*a*Th12	*a*Th13
		3.39			
		1.06	1.09	1.08	
	0.98		1.05		1.08
		1.82			
		1.10		1.04	0.99
		1.07			
1.21[b]		0.82			
	1.00	0.76		0.88	0.93
0.88			0.86		
1.00	2.00				
		0.98			
4	4	11	3	3	3
neutr.	neutr.	neutr.	neutr.	neutr.	neutr.

[b] 70 h hydrolysis gives two valines.

*a*Th20	*a*Th21	*a*Th22	*a*Th23	*a*Th24	*a*Th25	*a*Th26	*a*Th27
1.00	1.05						
		0.87	0.98		1.00		
	0.97		1.06				
	1.06	1.04				1.16	
	0.84						
		1.09	1.04				
2.00	1.14	1.12	1.03	2.10		1.96	1.05
						1.01	
	1.06			0.83			0.90
		0.87	0.91		2.00		
	0.91						
	0.96						
1.08	1.03	1.02	0.98	1.08	1.25	1.30	2.05
		0.64				1.03	
4	9	7	6	4	4	6	4
0.04	0.28	0.28	0.44	0.44	0.44	0.68	0.90

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

Peptide number	Amino acid sequence
*a*Th1	Val-(Ala,Tyr,Asp,Glu)
*a*Th2	Val-Ala-(Tyr,Asp,Glu,Leu,Ala)
*a*Th3	Phe-(Glu,Ala)
*a*Th4	Ala-Val-(Gly,Asp,Lys,Trp,Ser,Asp,Glu)
*a*Th5	Thr-(Glu,Lys,Gln,Asp,Ala)
*a*Th6	Phe-(Thr,Glu,Lys,Gln,Asp,Ala)
*a*Th7	Ala-(Val,Gly,Asp,Lys,Trp,Ser,Asp,Glu,Leu,Ser,Arg,Ala)
*a*Th8	Val-Val-Phe-Tyr
*a*Th9	Leu-Phe-Ser-Phe
*a*Th10	Leu-(Ala,Asn,Pro,Thr,Asp,Gly,Val,Asn,Pro,Lys)
*a*Th11	Tyr-Thr-Ser
*a*Th12	Leu-(Thr,Gly)
*a*Th13	Leu-Gly-Ser
*a*Th14	Leu-Val-(Arg,Asp,Ser,Ala,Gly,Gln)
*a*Th15	Thr-(Ser,Ile,Leu,Glu,Lys,Ala,Pro)
*a*Th16	Leu-Ala-Ala-Ala
*a*Th17	Val-(Val,Lys,Glu,Ala,Leu)
*a*Th18	Val-Ala
*a*Th19	Leu-(Phe,Ala,Leu),
*a*Th20	Ala-(Ala,Lys,Asp)
*a*Th21	Phe-Lys-Ala-Asn-Ile-Pro-Gln-Tyr-Ser
*a*Th22	Leu-(Thr,Gly,His,Ala,Glu,Lys)
*a*Th23	Leu-Lys-Ala-(Ser,Gly,Thr)
*a*Th24	Ile-(Lys-Ala-Ala)
*a*Th25	Leu-Leu-Lys-Thr
*a*Th26	Val-(His,Ala,Gln,Lys,Ala)
*a*Th27	Ile-Lys-(Ala,Lys)

Table 3. Amino acid sequence of peptide *a*Th21.

Sequence *a*Th21	Phe-Lys-Ala-Asn-Ile-Pro-Gln-Tyr-Ser					
	|—T2—|————————T1————————|					
Dansylation	DNS-Phe					
Tryptic peptides						
T1	Ala, 0.97;	Asp, 1.00;	Ile, 0.99;	Pro, 1.03;	Glu, 1.07;	
(neutral)	Tyr, 0.60;	Ser, 0.95				
Edman degradation						
Step 1	*Ala, 0.00*	Asp, 1.03;	Ile,0.98;	Pro, 0.97;	Glu, 1.10;	
	Tyr, 0,97;	Ser,0.95				
Step 2		*Asp, 0.21*	Ile, 0.91;	Pro, 1.18;	Glu; 1.04;	
	Tyr, 0.89;	Ser, 0.98				
Step 3			*Ile, 0,32*	Pro, 1.03;	Glu, 1.08;	
	Tyr, 0.89;	Ser, 1.01				
Step 4				*Pro, 0.63*	Glu, 1.07;	
	Tyr, 0.91;	Ser, 1.02				
Step 5					*Glu, 0.62*	
	Tyr, 0.83;	Ser, 1.02				
Carboxypeptidase A 15 min	Ser, 1.00;	Tyr,0.76				
T2 (E_{Lys}, 0.64)	Phe, 0.80;	Lys, 1.14				
Dansylation	DNS-Phe					

Peptide a*Th23*. Leucine was found to be the NH_2-terminal amino acid residue. Digestion of *a*Th23 with trypsin 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, *Ala-(Ser,Gly,Thr)*, is compatible with the NH_2-terminal sequence of the tryptic peptide *a*T6. The sequence of T2 was found to be *Leu-Lys*, which represents the COOH-terminal sequence of peptide *a*T9 as well as that of *a*T10. Since *a*T10 is linked to *a*T5 (peptide *a*Th25), *a*Th23 provides the overlap linking *a*T9 to *a*T6.

Table 4. Amino acid sequences of peptides *a*Th23 and *a*Th25.

Sequence *a*Th23	Leu-Lys-Ala-[Ser,Gly,Thr]
	|—T2—|————T1————|
Dansylation	DNS-Leu
Tryptic peptides	
T1	Ala, 1.05; Ser, 1.00; Gly, 0.93; Thr, 1.02
(neutral)	
Dansylation	DNS-Ala
T2	Leu, 1.07; Lys, 0.93
(E_{Lys}, 0.70)	
Dansylation	DNS-Leu
Sequence *a*Th25	Leu-Leu-Lys-Thr
	|——T2————|—T1-
Dansylation	DNS-Leu
Tryptic peptides	
T1	Thr, 1.00
(neutral)	
T2	Leu, 1.98; Lys, 1.02
(E_{Lys}, 0.58)	

Peptide a*Th25*. Leucine was established as the NH_2-terminal amino acid residue (Table 4). Tryptic digestion of *a*Th25 yielded two fragments, T1 and T2, which were electrophoretically separated at pH 6.5. T1 was found to be threonine, the NH_2-terminus of the tryptic peptide *a*T5. The sequence of T2 is concluded to be *Leu-Leu-Lys*, the unique NH_2-terminal sequence of the tryptic peptide *a*T10. *a*Th25, therefore, provides the overlap joining *a*T10 to *a*T5.

Because the sequences of the tryptic peptides of Lb*a* have already been established the sequences of the thermolytic peptides could be evaluated from their amino acid composition and from their NH_2-terminal analyses. Consequently, we were able to decide in an unequivocal fashion which tryptic peptides were connected to each other in the original Lb*a* chain (Fig. 1).

These results confirm the order of the tryptic peptides which was deduced from the chymotryptic peptides, and also show that no fragment was neglected during the isolation of tryptic peptides.

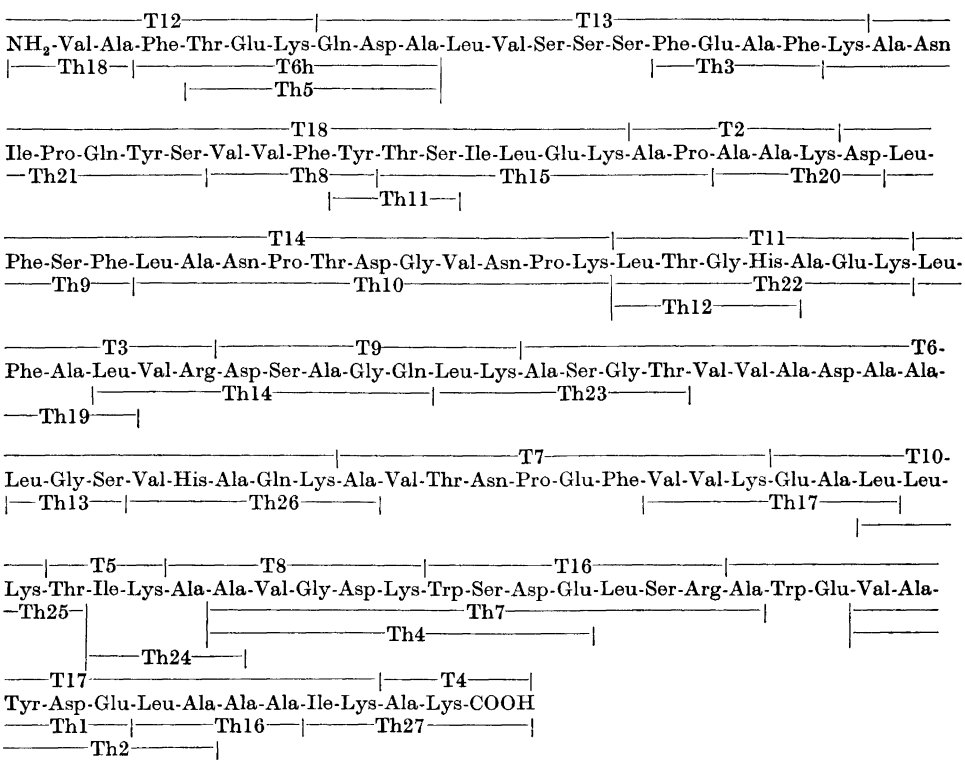

Fig. 1. The amino acid sequence of soybean leghemoglobin *a*. The tryptic peptides (T) are shown above the sequence and the thermolytic peptides (Th) below it.

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.* **27** (1973) 3817.
2. Ellfolk, N. and Sievers, G. *Acta Chem. Scand.* **26** (1972) 1155.
3. Ellfolk, N. and Sievers, G. *Acta Chem. Scand.* **27** (1973) 3371.
4. Dayhoff, M. O., Ed., *Atlas of Protein Sequence and Structure* **5** (1972).
5. Ellfolk, N. *Acta Chem. Scand.* **14** (1960) 609.
6. Ellfolk, N. *Acta Chem. Scand.* **16** (1962) 831.
7. Offord, R. E. *Nature* **211** (1966) 591.

Received June 28, 1973.