ACTA CHEMICA SCANDINAVICA 25 (1971) 3532–3548

Short Communications

The Primary Structure of Soybean Leghemoglobin

NILS ELLFOLK and GUNNEL SIEVERS

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

The amino acid sequences of two histidine-containing peptides obtained by tryptic hydrolysis of the apoprotein of the slow component of soybean leghemoglobin (Lb*a*) have been reported earlier.[1,2] We give here a preliminary description of the primary structure of the apoprotein of the same component (Lb*a*).

Materials and methods The slow component of soybean leghemoglobin (Lb*a*) and its apoprotein were prepared as described previously.[1] Hydrolysis of the apoprotein by trypsin, chymotrypsin, and thermolysin, purification of the peptide fragments and determination of their sequences will be published elsewhere.[3] Eighteen tryptic, twelve chymotryptic and fifteen thermolytic peptides were obtained. The peptides were purified by ion exchange chromatography, gel filtration, paper chromatography, and paper electrophoresis. The amino acid sequences of the individual peptides were determined using a variety of standard techniques, such as the Edman method, leucineaminopeptidase and carboxypeptidase A digestion. Some of the peptides were hydrolyzed further with thermolysin or substilopeptidase A and the sequences of the resulting small peptides were determined as above.

Results and discussion The overlapping peptides and the sequence data derived from tryptic, chymotryptic, and thermolytic peptides revealed the complete amino acid sequence of the apoprotein of the slow component of soybean leghemoglobin (Fig. 1). The amino acid composition of the ordinary Lb*a* chain deduced from the sequence is as follows: Lys_{14} His_2 Arg_2 Asp_{12} Thr_7 Ser_{11} Glu_{13} Pro_5 Gly_6 Ala_{26} Val_{14} Ile_4 Leu_{13} Tyr_3 Phe_3 Try_2. The total of 142

1 5 [Glu-Asn] 10 [Tyr] [Tyr] 20
NH_2-Val-Ala-Phe-Thr-Glu-Lys-Gln-Asp-Ala-Leu-Val-Ser-Ser-Ser-Phe-Glu-Ala-Phe-Lys-Ala-Asn-
25 30 35 40
Ile-Pro-Gln-Tyr-Ser-Val-Val-Phe-Tyr-Thr-Ser-Ile-Leu-Glu-Lys-Ala-Pro-Ala-Ala-Lys-Asp-Leu-
45 50 55 60 65
Phe-Ser-Phe-Leu-Ala-Asn-Pro-Thr-Asp-Gly-Val-Asn-Pro-Lys-Leu-Thr-Gly-His-Ala-Glu-Lys-Leu-
70 75 80 85
Phe-Ala-Leu-Val-Arg-Asp-Ser-Ala-Gly-Gln-Leu-Lys-Ala-Ser-Gly-Thr-Val-Val-Ala-Asp-Ala-Ala-
90 95 100 105
Leu-Gly-Ser-Val-His-Ala-Gln-Lys-Ala-Val-Thr-Asn-Pro-Glu-Phe-Val-Val-Lys-Glu-Ala-Leu-Leu-
110 115 120 125 130
Lys-Thr-Ile-Lys-Ala-Ala-Val-Gly-Asp-Lys-Trp-Ser-Asp-Glu-Leu-Ser-Arg-Ala-Trp-Glu-Val-Ala-
135 140
Tyr-Asp-Glu-Leu-Ala-Ala-Ala-Ile-Lys-Ala-Lys-COOH

Fig. 1. The primary structure of soybean leghemoglobin *a*.

	55			60			65						
Soybean Lb*a*	-Val-	Asn-	Pro-	Lys-	Leu-	Thr-	Gly-	His-	Ala-	Glu-	Lys-	Leu-	Phe-
Human γ-chain	-Gly-	Asn-	Pro-	Lys-	Val-	Lys-	Ala-	His-	Gly-	Lys-	Lys-	Val-	Leu-

	70			75			80			85			
Lb*a*	-Ala-	Leu-	Thr-	Asn-	Ala-	Val-	Ala-	His-	Val-	Asp-	Asp-	Met-	Pro-
γ-Chain	-Leu-	Thr-	Ala-	Val-	Val-	His-	Val-	Thr-	Ser-	Leu-	Thr-	Ser-	Lys-

	90			95			100						
Lb*a*	-Asn-	Ala-	Leu-	Ser-	Ala-	Leu-	Ser-	Asp-	Leu-	His-	Ala-	His-	Lys-
γ-Chain	-Tyr-	Arg-	Leu-	Asp-	Val-	Asp-	Pro-	Glu-	Asn-	Phe-	Arg-	Leu-	Leu-

Fig. 2. Comparison between the environment of the heme-binding histidines in the soybean leghemoglobin *a* and in the human γ-chain.

amino acid residues corresponds to a molecular weight of 15 159 and 15 775 for the apoprotein and the intact hemoprotein, respectively. The iron content of Lb*a* was calculated to be 0.354 % (dry weight).

The single pairs of histidine residues in Lb*a*, which may co-ordinate to the iron atom in the heme group as the fifth and sixth ligands, are in positions 61 and 92. It is interesting to note that one of the potential histidine ligands in hemoglobin and myoglobin [4] is located near to the center of the peptide chain. No -Lys-Lys-, -Arg-Lys- or -Lys-Arg- sequences are found in Lb*a*, in contrast to other hemoglobins, a characteristic feature of which is the sequence -Lys-Lys- situated close to one of the heme-binding histidine residues forming the characteristics "basic center" of the chains. The spacing between the two histidine residues in Lb*a* is 30 amino acid residues as compared to 28 residue spacing between the histidine ligands in the α- and β-chains of vertebrate hemoglobins and myoglobins.[5] This invariancy suggests that these residues are important for the maintenance of the structure of the protein molecule and thereby for its function.

The two tyrosine residues in parenthesis represent a peptide which was isolated from one leghemoglobin preparation only (obtained from soybean cultivated in the field). It seems that under certain conditions the slow component of leghemoglobin possesses two types of *a* chain. The ordinary type has phenylalanine residues in positions 15 and 18 while the other has tyrosine residues at these locations. Using the amino acid code given by Crick [6] the nucleotide triplets coding for the amino acids in positions 15 and 18 are $UU\overset{U}{C}$ and $UA\overset{U}{C}$ for phenylalanine and tyrosine, respectively, differing by only one base. This means that the minimum number of mutations required for conversion of this phenylalanine residue to a tyrosine residue is one. It seems that the variant tyrosines are associated with a deaminated glutamate in position 7 and an amidated aspartate in position 8. However, because of the lack of material there is some uncertainty in the sequence of this peptide, which represents an only occasionally formed minor component.

In Fig. 2 a comparison is made between the environment of the heme-binding histidine residues in Lb*a* and in the γ-chain of human hemoglobin.[7] Of these 48 amino acids 12 are identical. The comparison of the two chains justifies the assumption that they are homologous peptide chains, which means that leghemoglobin, like all hemoglobins, has evolved from the same primitive ancestor. If hemoglobins are assumed to be compounds characteristic of animals, it is possible that leghemoglobin represents a compound whose genetic information has been preserved in the leguminous plants from a time when no sharp difference existed between the plant and animal kingdoms.

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

1. Ellfolk, N. and Sievers, G. *Acta Chem. Scand.* **23** (1969) 2994.
2. Aggarwal, S. J. and Riggs, A. *Acta Chem. Scand.* **24** (1970) 2234.
3. Ellfolk, N. and Sievers, G. *Acta Chem. Scand. In press.*
4. Perutz, M. F. *J. Mol. Biol.* **13** (1965) 646.

5. Dayhoff, M. O., Ed., *Atlas of Protein Sequence and Structure* 4 (1969).
6. Crick, F. H. C. *Cold Spring Harbor Symp. Quant. Biol.* 31 (1966) 1.
7. Schroeder, W. A., Shelton, J. R., Shelton, J. B., Cormick, J. and Jones, R. T. *Biochemistry* 2 (1963) 992.

Received October 21, 1971.

Reindarstellung von Alkalicyanaten

ERIK SÖDERBÄCK

Chemisches Institut der Universität Uppsala, Box 531, S-751 21 Uppsala 1, Schweden

Handelspräparate von Natrium- und Kaliumcyanat reagieren infolge Verunreinigung durch Karbonat immer alkalisch. Bei Versuchen, neutral reagierende, karbonatfreie Präparate der Cyanate herzustellen, müssen nach Ansicht des Verfassers Wasser und Alkohole als Lösungsmittel vermieden werden; anwendbar ist dagegen wasserfreies Aceton, worin die Alkalicyanate sehr wenig löslich sind. Als Ausgangsmaterial dient Merkuricyanat, $Hg(OCN)_2$, das nach Söderbäck [1] aus Silbercyanat und $HgCl_2$ in Äther dargestellt wird. Wenn eine Acetonlösung des Merkuricyanats mit einer Acetonlösung von NaI, das ebenfalls in Aceton leicht löslich ist, zusammengebracht wird, fällt sofort ein weisser Niederschlag von NaOCN aus:

$$Hg(OCN)_2 + 3NaI \rightarrow 2NaOCN + NaHgI_3$$

Da $NaHgI_3$ in Aceton leicht löslich ist, braucht man nur den Niederschlag mit Aceton zu waschen und in Vacuum zu trocknen um reines NaOCN zu erhalten; die Ausbeute ist quantitativ.

Bei der analogen Darstellung von KOCN wird das in Aceton spärlich lösliche KI durch das löslichere KSCN ersetzt. Von Komplexionen, die $Hg(SCN)_2$ mit SCN^- in Wasser gibt, sind $Hg(SCN)_3^-$ und $Hg(SCN)_4^{2-}$ beschrieben worden; über die Komplexbildung in wasserfreiem Aceton ist nichts bekannt. Wenn man aber die für eine Bildung von $Hg(SCN)_4^{2-}$ nötige Menge von KSCN anwendet, erhält man das schwerlösliche KOCN in quantitativer Ausbeute, da die Komplexsalze $KHg(SCN)_3$ und $K_2Hg(SCN)_4$ in Aceton leicht löslich sind. Analog kann NaOCN aus $Hg(OCN)_2$ und NaSCN dargestellt werden. Um eine quantitative Ausbeute zu erhalten, muss man aber einen ziemlich grossen Überschuss von NaSCN anwenden, wenigstens 5 Mol NaSCN/Mol $Hg(OCN)_2$.

Es ist zu bemerken, dass $Hg(OCN)_2$ nicht unbegrenzt haltbar ist sondern allmählich in eine in Aceton und Äther unlösliche Substanz umgewandelt wird. Um die Umwandlung zu verzögern muss man das Salz bei niedriger Temperatur aufbewahren.

Experimentelles. Darstellung von NaOCN. (a) Aus $Hg(OCN)_2$ und NaI. 14,0 g NaI wurden in 50 ml Aceton gelöst und mit einer Lösung von 8,52 g $Hg(OCN)_2$ in 25 ml Aceton gemischt. Unter starker Wärmeentwicklung wurde weisses NaOCN ausgefällt. Nach einigen Stunden bei Zimmertemperatur wurde auf 0° abgekühlt, der Niederschlag auf einem gewogenen Glasfilter gesammelt, mit Aceton bis zum Verschwinden der I^- und Hg^{2+}-Reaktionen gewaschen und in Vacuum getrocknet. Ausbeute 3,90 g; theor. 3,90 g. Reaktion neutral.

(b) aus $Hg(OCN)_2$ und NaSCN. 5,68 g $Hg(OCN)_2$ in 15 ml Aceton + 8,30 g NaSCN in 50 ml Aceton. Verfahren genau wie oben. Erhalten: 2,59 g NaOCN; theor. 2,60 g.

Darstellung von KOCN aus $Hg(OCN)_2$ und KSCN. 8,0 g KSCN wurden in 90 ml Aceton gelöst und mit einer Lösung von 5,68 g $Hg(OCN)_2$ in 15 ml Aceton gemischt. Unter merkbarer Wärmeentwicklung wurde KOCN ausgefällt. Der Niederschlag wurde wie oben beschrieben behandelt. Ausbeute 3,23 g; theor. 3,24 g. Reaktion neutral.

1. Söderbäck, E. *Acta Chem. Scand.* 11 (1957) 1624.

Eingegangen am 30. Oktober 1971.