PENGUIN HEMOGLOBIN (*APTENODYTES FORSTERI*). A 45 RESIDUE *N*-TERMINAL SEQUENCE

C. MONIER*, A.G. SCHNEK**, J. DIRKX* and J. LEONIS

Laboratoire de Chimie Générale I, Faculté des Sciences, Université libre de Bruxelles, 1050 Bruxelles, Belgique

Received 29 June 1973

1. Introduction

The hemoglobins have been studied from the point of view of their amino acid sequence principally in mammals; this study has also been extended to the fish (carp, lamprey) and to the invertebrates (*Chironomus thummi, Glycera dibranchiata*). Nevertheless, as far as birds are concerned, only the primary structure of the α-chain of the chicken major hemoglobin component is known at the present time. With the goal of expanding the bulk of knowledge with respect to structural comparative studies, the determination of the amino acid sequence of the α-chain of the penguin hemoglobin was undertaken.

The positions of the first 45 residues of this α-chain were determined by two different approaches: on the one hand from information obtained on the tryptic peptides of the two cyanogen bromide fragments of the α-chain, and on the other hand from results of the degradation of the entire polypeptide chain by use of the sequencer.

2. Materials and methods

The hemoglobin was extracted from erythrocytes by hemolysis as was previously described [1]. After extraction of the heme by HCl acidified acetone [2], the globin was successively submitted to reduction by mercaptoethanol and to aminoethylation of the thiol groups by use of ethyleneimine [3]. The α- and β-chains were isolated from the aminoethylated globin [4].

The α-chain dissolved as per 1 mg per ml of 70% formic acid, was hydrolyzed by cyanogen bromide in excess as 100 moles per mole of protein. The fragments obtained by cleavage were fractionated by filtration on Sephadex G-25 fine column in 0.2 M acetic acid.

For the separated fragments the *N*-terminal residues were determined as dansylamino acids [5], whereas the *C*-terminal residues were identified by amino acid analysis after carboxypeptidases A or B hydrolysis.

The isolated fragments were digested by trypsin (pH 8.0, 37°C, 4 hr) with an enzyme/substrate weight ration of 1:50. The separation of tryptic peptides was realized either by high voltage paper electrophoresis at pH 3.7 [6], or by Dowex 50 X-2 chromatography [7].

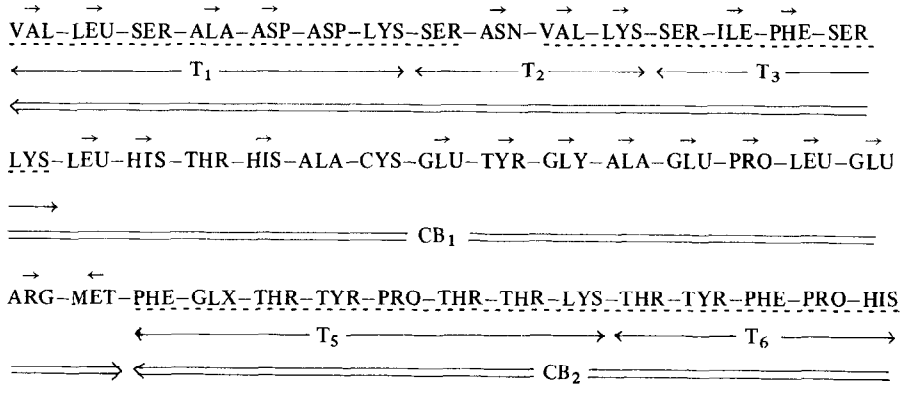
The amino acid analysis of the fragments obtained after use of cyanogen bromide and that of the tryptic peptides was realized with an automatic amino acid analyser (Beckman Unichrom).

The sequence of the peptides was determined by the dansyl-Edman technique [5, 8]. The chromatography methods of Woods and Wang [9] permitted the identification of dansyl amino acids.

The α-chain of the aminoethylated globin was submitted to Edman automatic degradation [10] with a Beckman sequencer (model 890). The PTH amino acids were identified by gas chromatography (Beckman G C 45 chromatograph) [11] or by thin-layer chromatography [12].

* From Service de Chimie Biologique, Faculté des Sciences, Université de l'Etat, 7000 Mons, Belgique.
** To whom to address correspondence.

Table 1
Determination of the *N*-terminal sequence of penguin α-chain hemoglobin.



← T → tryptic peptides; ⇐ CB ⇒ peptides from cyanogen bromide cleavage; → Amino acid identified without ambiguity by automatic sequential degradation; ← Amino acids identified by carboxypeptidases; – – – Amino acids identified as dansyl derivatives.

3. Results and discussion

The *N*-terminal sequence of the α-chain of penguin hemoglobin is indicated in table 1. The positions of the amino acids 1–32 were determined from 3 tryptic peptides obtained from the *N*-terminal fragment of the chain cleaved at the level of methionine in position 32 by cyanogen bromide, and by the automatic degradation realized by the use of the Edman technique on the sequencer. The amino acids 33–45 were determined from the sequence of 2 tryptic peptides obtained from the second cyanogen bromide fragment; they were replaced by homology with the α-chain of human hemoglobin, that part of the peptide chain being relatively invariant with respect to the different known α-chains.

In table 2, are underlined the replacements observed when one compares the first 45 residues of the penguin or the chicken α-chain [13] to the corresponding amino acid's sequence of the human one [14, 15].

The N-terminal parts of the α-chains seemed therefore relatively variable (13–15 substitutions on 45 residues).

One finds for the α-chains of the hemoglobins of the penguin and chicken the presence in position 13

Table 2
Alignment of the first 45 residue sequences of the human, chicken and penguin hemoglobin α-chains.

	1	5	10	15
α Human	VAL–LEU–SER–PRO–ALA–ASP– LYS–THR–ASN–VAL–LYS–ALA–ALA–TRY–GLY–LYS–VAL–GLY–			
α Chicken	VAL–LEU–SER–ASN–ALA–ASP–LYS–ASN–ASN–VAL–LYS–GLY–ILE–PHE–THR–LYS–ILE–ALA–			
α Penguin	VAL–LEU–SER–ALA–ASP–ASP–LYS–SER–ASN–VAL–LYS–SER–ILE–PHE–SER–LYS–LEU–HIS			

	20	25	30
α Human	ALA–HIS–ALA–GLY–GLU–TYR–GLY–ALA–GLU–ALA–LEU–GLU–ARG–MET–PHE–LEU		
α Chicken	GLY–HIS–ALA–GLU–GLU–TYR–GLY–ALA–GLU–THR–LEU–GLU–ARG–MET–PHE–ILE		
α Penguin	THR–HIS–ALA–CYS–GLU–TYR–GLY–ALA–GLU–PRO–LEU–GLU–ARG–MET–PHE–GLX		

	35	40	45
α Human	SER–PHE–PRO–THR–THR–LYS–THR–TYR–PHE–PRO–HIS		
α Chicken	GLY–PHE–PRO–THR–THR–LYS–THR–TYR–PHE–PRO–HIS		
α Penguin	THR–TYR–PRO–THR–THR–LYS–THR–TYR–PHE–PRO–HIS		

and 14, of Ile and Phe whereas for the other known α-chains one finds Ala and Tyr.

In the 45 positioned residues, one finds in position 43 a phenylalanine which is invariant for all different known hemoglobins and myoglobins.

A more complete study of the total sequence of the α-chain of the penguin hemoglobin will permit more precise phylogenetic conclusions.

References

[1] Schnek, A.G., Paul, C. and Leonis, J. (1966) International Symposium on comparative Haemoglobin Structure, Thessaloniki, p. 103.
[2] Rossi-Fanelli, A., Antonini, E. and Caputo, A. (1968) Biochim. Biophys. Acta 30, 608.
[3] Crestfield, A.M., Moore, S. and Stein, W.H. (1963) J. Biol. Chem. 288, 622.
[4] Paul, C., Monier, C., Schnek, A.G., Brygier, J. and Leonis, J. (1968) Arch. Internatl. Physiol. Biochem. 76, 383.
[5] Gray, W.R. and Hartley, B.S. (1963) Biochem. J. 89, 59.
[6] Katz, A.M., Dreyer, W.S. and Anfinsen, C.B. (1959) J. Biol. Chem 234, 2897.
[7] Schroeder, W.A., Jones, R.T., Cornick, J. and McCalla, K. (1962) Anal. Biochem. 34, 1570.
[8] Dopheide, T.A.A., Moore, S. and Stein, W.H. (1967) J. Biol. Chem. 242, 1833.
[9] Woods, K.R. and Wang, K.T. (1967) Biochim. Biophys. Acta 133, 369.
[10] Edman, P. and Begg, C. (1967) European J. Biochem. 1, 80.
[11] Pisano, J.J. and Bronzert, T.J. (1969) J. Biol. Chem. 244, 5597.
[12] Edman, P. (1970) in: Protein Sequence Determination, (Needleman, S.B., ed.), p. 211, Springer-Verlag, Berlin, Heidelberg, New York.
[13] Matsuda, G., Takei, H., Ku Chayu, Wu and Shiozowa, T. (1971) Internatl. J. Protein Res. 3, 173.
[14] Braunitzer, G., Gehring-Muller, R., Hilschmann, N., Hilse, K., Hobam, G., Rudloff, V. and Wittman-Liebold, B. (1961) Hoppe Seyler's Z. Physiol. Chem. 325, 283.
[15] Hill, R.J., Konigsberg, W., Guidotti, G. and Craig, L.C. (1962) J. Biol. Chem. 237, 1549.