PHYLOGENY OF HEMOGLOBINS: AMINO ACID SEQUENCE OF THE N-TERMINAL PART OF A VIPER (*VIPERA ASPIS*) HEMOGLOBIN α-CHAIN

Michel DUGUET, Jean-Pierre CHAUVET and Roger ACHER
Laboratory of Biological Chemistry, 96 Boulevard Raspail, Paris VI, France

Received 8 November 1972

1. Introduction

In a previous paper purification of viper hemoglobin as well as separation of the polypeptide chains by counter-current distribution of carboxymethylated globin have been described [1]. Component A was identified as an α-type chain [1]. Sequence analysis of this protein has been undertaken in order to compare a reptilian hemoglobin α-chain with the known mammalian α-chains on one hand and those of chicken and carp on the other [2]. This paper describes the amino acid sequence of the N-terminal part of the viper α-chain.

2. Results

2.1. *Preparation of an N-terminal fragment of α-chain (CNI)*

Because component A is the only chain of viper hemoglobin having a methionine residue [1], it was convenient to prepare an N-terminal fragment of this chain by treating directly the whole globin with cyanogen bromide [3] and isolating the fragment by gel filtration. The yield of cleavage turned out to be higher with globin than with the purified α-chain, probably because some modification of methionine occurs during counter-current distribution.

Globin (1100 mg), prepared as previously described [1], was dissolved in 70% formic acid (20 ml) and treated with cyanogen bromide (1 g) for 24 hr. The solution was then diluted with water (80 ml) and freeze-dried. The N-terminal fragment of α-chain was separated by gel filtration; samples of 100 mg of material were passed through a 4.5 × 200 cm column of Sephadex G-50 equilibrated with 0.5 M acetic acid and fractions of 10 ml were collected. Fractions containing the N-terminal fragment (tubes 110 to 125) were pooled and after further purification on Sephadex G-25, 35 mg (about 10 μmole) of polypeptide were obtained from about 1 g of globin.

2.2. *Preparation of tryptic peptides from the N-terminal fragment*

Amino acid analysis of the N-terminal fragment, determined with samples of 10–20 nmole by using a Spinco 120 B automatic analyzer [4], indicated that the polypeptide contains 32 residues, two of which are lysine and three arginine [1]. Tryptic cleavage (weight ratio enzyme/substrate 1%, 37°, 0.1 M ammonium bicarbonate, pH 8.0, 3 hr) produced 5 peptides and homoserine. The amino acid sequence of four of them (T_1 to T_4) was previously established by using Edman degradation [5, 6] and the alignment of these four peptides was determined through maleylated overlapping peptides [1]. The fifth tryptic peptide (T_5), not yet sequenced, contains 15 residues and occupies the C-terminal position in CNI just before homoserine [1].

Preparation of rather large amounts (3–4 μmoles) of peptide T_5 was carried out by cleaving the citraconylated [7] polypeptide CNI with trypsin, isolating the C-terminal fragment, unblocking the lysine residue and cleaving again with trypsin. 35 mg (10 μmoles) of CNI were dissolved in 3 ml of water and the pH was adjusted at 9.0 with NH_4OH. Six additions of 10 μl of citraconic anhydride were made at 20 min intervals, the pH being maintained between 8.2 and

 North-Holland Publishing Company – Amsterdam

8.3 with 5 N NaOH. The mixture was then kept for 90 min at room temp. and reagents were removed by gel filtration on Sephadex G-25 in 0.1 M ammonium bicarbonate. The citraconylated peptide was cleaved by trypsin under the conditions described above. Three peptides and homoserine were obtained and separated by gel filtration on Sephadex G-25 equilibrated with 0.1 M ammonium bicarbonate. Peptide $CNICT_1$ (7 mg; about 6 μmoles) is similar by amino acid composition and N-terminal sequence (Val–Leu–Ser–Glu–Asp) to the maleylated peptide MT_1 previously characterized and which accounts for the first two tryptic peptides of α-chain, T_1 and T_2 ([1] and table 1). $CNICT_{II}$ is also similar to MT_{II} shown to be a dipeptide Val–Arg (T_3). $CNICT_{III}$ (14 mg; 6 μmoles) corresponds to MT_{III} which is an overlapping peptide containing two tryptic peptides T_4 and T_5. The study was therefore concentrated on peptide $CNICT_{III}$.

Removal of citraconyl groups was carried out in 0.2 M acetic acid (pH 3.25, 6 hr, room temp.) [7], and the peptide was subjected to tryptic hydrolysis under the usual conditions. Because it contains a single lysine residue, tryptic cleavage gave two fragments which were separated by gel filtration on a column (2.3 × 60 cm) of Sephadex G-25 equilibrated with 0.1 M ammonium bicarbonate. After further purification on the same column, 1.5 mg (about 3 μmoles) of the smaller fragment was obtained. Amino acid analysis showed that this peptide was identical to the tryptic peptide T_4 previously characterized and Edman degradation confirmed the sequence Thr–Ser–Val–Gly–Lys [1]. 5 mg (3 μmoles) of the larger fragment was purified under the same conditions. It had 15 residues (Asp_1, Thr_2, Ser_1, Glu_3, Gly_2, Pro_2, Leu_2, Tyr_1, Arg_1) and Edman degradation gave the N-terminal sequence Asn–Pro–Glu. This peptide was therefore similar by amino acid composition and N-terminal residue to the tryptic peptide T_5.

2.3. *Determination of the amino acid sequence of peptide T_5*

4 mg of peptide T_5 (about 2.5 μmoles) were subjected to chymotryptic hydrolysis (weight ratio enzyme/substrate 1%; 0.1 M ammonium bicarbonate pH 8.0; 37°; 2 hr). The material was freeze-dried and the mixture of peptides was examined by analytical chromatoelectrophoresis. Four peptides were detected with ninhydrin; specific staining reactions [8] revealed that two of them contained arginine and one had tyrosine. Preparative separation of peptides has been carried out by ion-exchange chromatography on a column (1 × 20 cm) of Dowex-50-X_2 with a linear gradient of pyridine acetate (0.2 M pH 3.1 to 2.0 M pH 5.0; 440 ml with a flow rate 30 ml/hr; 40°). Four peaks were detected by alkaline hydrolysis and ninhydrin reaction. The fractions Ch_1, Ch_2, Ch_3 and Ch_4 were collected and freeze-dried. Further purification was performed by paper chromatography and 0.4 to 1.2 μmole of each chymotryptic peptide was finally obtained. About 1/20 was used for amino acid analysis, 4/20 for carboxypeptidase A or B hydrolysis [9] (weight ratio enzyme/substrate 2%; 0.1 M ammonium bicarbonate pH 8.0; 37°; 45 min) and the remainder for Edman degradation (table 1). T_5Ch_1 was obviously the N-terminal fragment because it had the N-terminal sequence Asn–Pro–Glu of T_5 and T_5Ch_4 was the C-terminal one because it contained arginine. The sequence of peptide T_5 could therefore be deduced (table 1).

2.4. *Sequence of the first 34 residues of the α-chain*

The amino acid sequence of the first 16 residues of the α-chain has been confirmed. The alignment of tryptic peptides T_4–T_5 is given by peptide $CNICT_{III}$ and free homoserine is obviously the C-terminal residue of the fragment CNI so that the amino acid sequence of the first 32 residues can be determined (table 1). Cyanogen bromide cleaves α-chain into two fragments, an N-terminal peptide and a large polypeptide of about 110 residues. Edman degradation showed that the N-terminal sequence of this large fragment was Phe–Ala. The sequence of the first 34 residues of α-chain could therefore be deduced from all the data (tables 1 and 2).

3. Discussion

When the sequence of the first 34 residues of the viper hemoglobin α-chain is compared to the homologous part of the 22 mammalian, the chicken and the carp α-chains which are known to date [2], one can observe that amino acids in positions 2, 3, 6, 7, 16, 27 and 31 are invariant. Furthermore, in all the 24 non-reptilian proteins, the residue in position 11

Table 1
Peptides of the N-terminal fragment CNI.

Peptide	Sequence	Number of residues
	1 7 9	
CT_I (T_1–T_2)	Val–Leu–Ser–Glu–Asp–Asp–Lys–Asn–Arg	9
	10 11	
CT_{II} (T_3)	Val–Arg	2
	12 16	
T_4	Thr–Ser–Val–Gly–*Lys*	5
	17 24	
T_5Ch_1	Asn–Pro–Glu–Leu–Pro–Gly–Glu–Tyr	8
CT_{III} T_5Ch_2	Gly–Ser–Glu–Thr–Leu	5
T_5Ch_3	Thr–Arg	2
	25 31	
T_5Ch_4	Gly–Ser–Glu–Thr–Leu–Thr–Arg	
	32	
CT_{IV}	*Hse*	1

→Determination by Edman degradation [5, 6].←Determination by carboxypeptidase A or B [9]. CT: Tryptic fragments of citraconylated CNI. T: Tryptic peptides of CNI. T_5Ch: Chymotryptic peptides of T_5.

Table 2
Comparison of hemoglobin α-chains

	Species	Million years since the divergence	Sequence of the first 34 residues
			1 5 10 15
Mammals	Man	–	Val–Leu–Ser–Pro–Ala–Asp–Lys–Thr–Asn–Val–Lys–Ala–Ala–Trp–Gly–
Mammals	Ox	80	– – – Ala – – – Gly – – – – – – –
Mammals	Gray kangaroo	130	– – – Ala – – – Gly–His – – – Ile – –
Birds	Chicken	300	– – – Asn – – – Asn – – – Gly–Ile–Phe–Thr
Reptiles	Viper	300	– – – Glu–Asp – – Asn–Arg – Arg–Thr–Ser–Val –
Fish	Carp	450–500	Ac-Ser – – Asp–Lys – – Ala–Ala – – Ile – – Ala

The dashes represent residues identical with those of human α-chain.

is lysine which is replaced by arginine (a conservative substitution) in viper α-chain. Of 7 invariant residues, five are charged residues (Asp-6, Lys-7, Lys-16, Glu-27 and Arg-31) and are probably involved in salt bridges necessary for the stability of the molecule.

Table 2 shows the substitutions in ox, kangaroo, chicken, viper and carp α-chains with human α-chain taken as reference. The numbers of substitutions are 4, 9, 12, 17 and 17, respectively, for approx. 80, 130, 300, 300 and 450 million years of divergence since the common ancestor. Although the comparison is limited to the first 34 residues of the polypeptide chain and to a few species, it is rather unexpected to find the same number of substitutions for a reptile and for a bony fish; moreover although the evolutionary lines leading respectively to modern mammals, birds and snakes have diverged some 300 million years ago, chicken α-chain seems significantly nearer the mammalian α-chains than the viper α-chain. On the other hand viper α-chain is as far from chicken α-chain as from human α-chain (17 substitutions in both cases). Either avian lines have diverged from mammalian lines more recently than assumed or, more likely, the rates of substitutions were not similar in the different lines.

Mammals, birds and reptiles seem to have a non-acetylated α-chain in contrast to amphibians [10, 11] and fish [12]. It is noteworthy that when α-chain is not acetylated, the N-terminal residue is always valine but short side-chain residues such as serine and alanine are found for acetylated α-chains.

The first 35 amino acids of α-chain are arranged in two helical segments A (residues 3–18) and B (residues 20–35) described by Perutz [13, 14]. Among remarkable changes there is in position 17 (A_{15}) a polar residue (Asn) in place of the usual non-polar (Val or Ile) located in the interior of the molecule [15]; in position 18 (A_{16}) there is proline in place of glycine found in mammals, and again a proline in position 21 (B_2) in place of alanine or glycine. The last turn of helix A might be distorted but not the first turn of helix B since proline can be tolerated at the beginning of an α-helix.

Three residues out of the 16 involved in the contacts between α_1- and β_1-chains [12], are in the N-terminal part of α-chain, namely residues no. 30, 31, and 34. Residues no. 30 (Thr) and 31 (Arg) are polar

of species from four vertebrate classes.

																											Number of substitutions
16							20										25							30								34		
Lys	–	Val	–	Gly	–	Ala	–	His	–	Ala	–	Gly	–	Glu	–	Tyr	–	Gly	–	Ala	–	Glu	–	Ala	–	Leu	–	Glu	–	Arg	–	Met	–	Phe	–	Leu	
–	–	–		Gly	–	–		Ala	–	–		–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–		4
–	–	–		Gly	–	–				–		–	–	Ala	–	–	–	Gly	–	–	–	–	–	–	–	–	–	Thr	–	–	–	His				9
–	Ile	–	Ala	–	Gly	–	–	Glu	–	–	–	–	–	–	–	–	–	Thr	–	–	–	–	–	–	–	–	–	–	–	–	–	Ile				12
–	Asn	–	Pro	–	Glu	–	Leu	–	Pro	–	–	–	–	–	–	Ser	–	Thr	–	–	–	Thr	–	–	–	–	–	–	–	–	–	Ala				17
–	Ile	–	Ser	–	Pro	–	Lys	–	–	Asp	–	Asp	–	Ile	–	–	–	–	–	–	–	Gly	–	–	–	Leu	–	Thr								17

and residue no. 34 (Ala) non-polar, as in most other known α-chains. Of the 10 amino acids involved in the contacts between α_1- and β_2-chains, none is located in the two helical segments A and B of α-chain [12]. Finally an only residue (Met-32) out of the 19 involved in the interactions between α-chain and heme is located in the 34-residue N-terminal sequence [12]. Because the contact residues are much less subjected to substitution than the others [6], the N-terminal sequence of the α-chain can be regarded as a particularly variable part of the protein.

Acknowledgement

The authors with to thank Miss Christiane Devaux for her skilled technical assistance.

References

[1] M. Duguet, J.P. Chauvet and R. Acher, FEBS Letters 18 (1971) 185.
[2] M.O. Dayhoff, Atlas of Protein Sequence and Structure, The National Biomedical Research Foundation, Vol. 5 (1972).
[3] J. Schreiter and W. Witkop, J. Am. Chem. Soc. 86 (1964) 2441.
[4] D.H. Spackman, W.H. Stein and S. Moore, Anal. Chem. 30 (1968) 1190.
[5] W.A. Schroeder, J.R. Shelton, J.B. Shelton, J. Cormick, and R.T. Jones, Biochemistry 2 (1963) 992.
[6] J. Chauvet and R. Acher, Biochemistry 11 (1972) 916.
[7] H.B.F. Dixon and R.N. Perham, Biochem. J. 109 (1968) 312.
[8] R. Acher and C. Crocker, Biochim. Biophys. Acta 9 (1952) 704.
[9] H. Fraenkel-Conrat, J.I. Harris and A.L. Levy, Methods Biochem. Anal. 2 (1955) 359.
[10] J.P. Chauvet and R. Acher, Intern. J. Protein Res. 3 (1971) 261.
[11] J.P. Caffin, J.P. Chauvet and R. Acher, FEBS Letters 5 (1969) 196.
[12] K. Hilse, V. Sorger and G. Braunitzer, Hoppe-Seyler's Z. Physiol. Chem. 344 (1966) 166.
[13] M.F. Perutz, H. Muirhead, J.M. Cox and L.C.G. Goaman, Nature, 219 (1968) 131.
[14] M.F. Perutz, Nature, 228 (1970) 726.
[15] M.F. Perutz, J.C. Kendrew and H.C. Watson, J. Mol. Biol. 13 (1965) 669.