PHYLOGENY OF HEMOGLOBINS: THE COMPLETE AMINO ACID SEQUENCE OF AN α-CHAIN OF VIPER (*VIPERA ASPIS*) HEMOGLOBIN

Michel DUGUET, Jean-Pierre CHAUVET and Roger ACHER
Laboratory of Biological Chemistry, 96 Bd Raspail, 75006 – Paris, France

Received 28 August 1974

1. Introduction

In two previous notes, the amino acid sequence of the first 92 residues of an α-chain of viper (*Vipera aspis*) hemoglobin has been described [1,2]. This paper deals with the C-terminal part and gives the complete amino acid sequence (141 residues) of viper α-chain. When compared to the human hemoglobin α-chain, the N-terminal portion of the molecule (helices A and B) appears more substituted than the middle (helices C, E and F) and C-terminal (helices G and H) parts; this suggests different rates of evolution for different parts of the same polypeptide chain and might be explained by the peculiar stability of residues involved in the contacts with heme or between sub-units.

2. Results

2.1. *Purification of peptides CN_{II} CT_{II} and CN_{II} CT_{III}*

Cleavage of carboxymethylated α-chain into two peptides, CN_I and CN_{II}, by cyanogen bromide [3] and purification of the fragments has been previously described [1,2]. The amino acid sequence of CN_I (32 residues) has been determined [1]; because CN_{II} (about 110 residues) has only 3 arginine residues, one of which being in the C-terminal position, cleavage of citraconylated CN_{II} by trypsin leads to three fragments CN_{II} CT_I (60 residues), CN_{II} CT_{II} (21 residues) and CN_{II} CT_{III} (28 residues). Purification and sequence analysis of CN_{II} CT_I has been described in a previous paper [2]. CN_{II} CT_{II} and CN_{II} CT_{III} are purified by gel filtration on a column of Sephadex G-50 fine (2.6 × 80 cm, 0.1 M ammonium bicarbonate). 4-ml fractions are collected; CN_{II} CT_{II} is recovered in fractions 58 to 72 and CN_{II} CT_{III} in fractions 48 to 58. The products are freeze-dried. From 120 mg of CN_{II}, 15 mg of CN_{II} CT_{II} and 20 mg of CN_{II} CT_{III} (approx. 6-7 μmol) are obtained.

Homogeneity is checked by Edman degradation carried out on 500 nmol of each peptide [4] and amino acid composition is determined with 8–15 nmol [5]. Fragment CN_{II} CT_{II} has an N-terminal sequence Val–Asp–Pro–Ala–Asn–Phe and fragment CN_{II} CT_{III} an N-terminal sequence Asn–Pro Glu–Phe–Gly–Pro–Ala. Because tyrosine is found only in CN_{II} CT_{III} and because tyrosine is the penultimate residue of the α-chain [6], it is clear that CN_{II} CT_{III} is the C-terminal fragment of CN_{II}, and that CN_{II} CT_{II} is located in the middle part of CN_{II}.

2.2. *Determination of the amino acid sequence of peptide CN_{II} CT_{II}*

Amino acid composition of CN_{II} CT_{II} shows that the peptide contains 21 residues, one of which is lysine and one arginine. Removal of citraconyl groups is carried out by acid treatment [7]; cleavage by trypsin is performed under previously described conditions [1,2] (enzyme substrate weight ratio 1% 37°C 0.1M ammonium bicarbonate, pH 8.0, 3 hr) on 8 mg of peptide (about 3–4 μmol). The material is freeze-dried and tryptic peptides are separated by peptide mapping [4]. Two peptides are detected with dilute ninhydrin, one of which giving a red color with Sakaguchi reagent [8]. They are eluted and analyzed. These two peptides correspond to tryptic units T_{13} and T_{14} obtained by direct tryptic hydrolysis of intact α-chain. T_{13} has 7 residues and

Table 1
Peptides of the fragment $CN_{II}\,CT_{II}$

Peptides	Sequence	Number of residues
	93 → → → → → → 113	
$CN_{II}\,CT_{II}$	Val–Asp–Pro–Ala–Asn–Phe . . . . . Arg	21
$CN_{II}\,CT_{II}$		
	93 → → → → → 99	
T_{13}	Val–Asp–Pro–Ala–Asn–Phe–Lys	7
	100 → → → → → → → 113	
T_{14}	Ile–Leu–Ser–Gln–Cys–Leu–Leu(Ser, Thr, Leu, Ala, Asn, His) Arg	14
		21
$CN_{II}\,CT_{II}$		
	93 → → → → 98	
Ch_1	Val–Asp–Pro–Ala–Asn–Phe	6
	99 → 101	
Ch_2	Lys–Ile–Leu	
	99 → → → → → 106	
Ch_{2-3}	Lys–Ile–Leu–Ser–Gln(Cys, Leu) Leu	8
	107 → 109	
Ch_4	Ser–Thr–Leu	3
	107 → → → → → 113	
Ch_{4-5}	Ser–Thr–Leu–Ala–Asn–His–Arg	
	110 → → 113	
Ch_5	Ala–Asn–His–Arg	4
		21

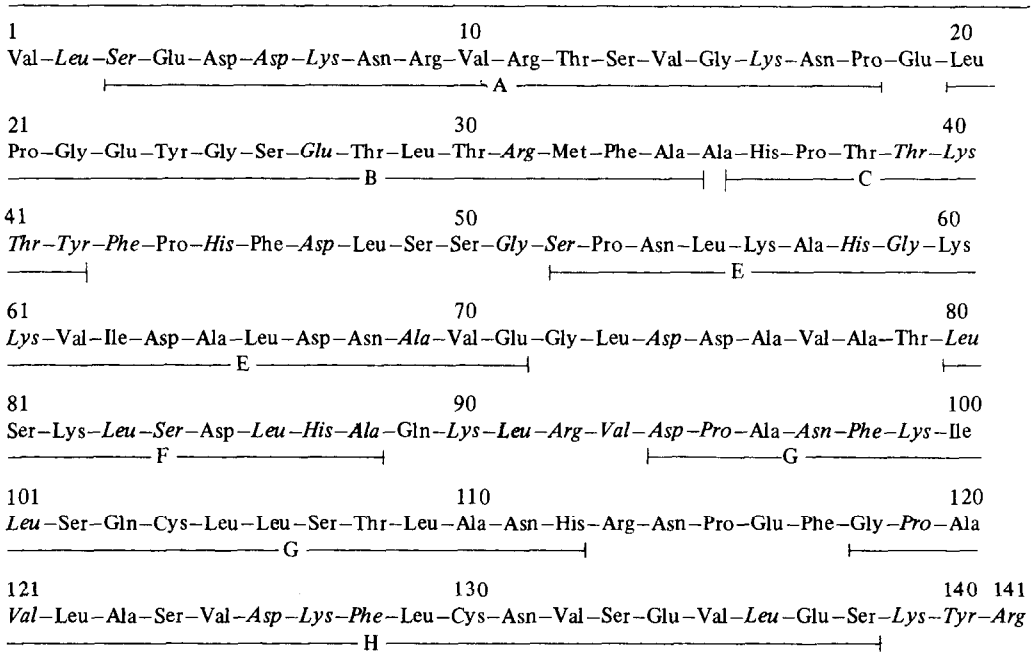
——→ Determination by Edman degradation [4] $CN_{II}\,CT_{II}$: tryptic fragment of citraconylated CN_{II}. T : tryptic peptides of the uncitraconylated $CN_{II}\,CT_{II}$, Ch : chymotryptic peptides of $CN_{II}\,CT_{II}$.

Table 2
Peptides of the fragment $CN_{II}\,CT_{III}$

Peptides	Sequence	Number of residues
	114 141	
$CN_{II}\,CT_{III}$	Asn–Pro–Glu–Phe–Gly–Pro–Ala . . . . . Arg	28
$CN_{II}\,CT_{III}$		
	114 127	
T_{15}	Asn–Pro–Glu–Phe–Gly(Pro, Ala, Val, Leu, Ala, Ser, Val, Asp) Lys	14
	128 139	
T_{16}	Phe–Leu–Cys–Asn–Val(Ser, Glu, Val, Leu, Glu, Ser) Lys	12
	140 141	
T_{17}	Tyr–Arg	2
		28
$CN_{II}\,CT_{III}$		
	114 117	
Ch_1	Asn–Pro–Glu–Phe	4
	118 122	
Ch_2	Gly–Pro–Ala–Val–Leu	5
	123 128	
Ch_3	Ala–Ser–Val–Asp–Lys–Phe	6
	118 128	
Ch_{2-3}	Gly–Pro–Ala–Val–Leu–Ala(Ser, Val, Asp, Lys) Phe	
	129 136	
Ch_4	Leu–Cys–Asn–Val(Ser, Glu, Val) Leu	8
	137 140	
Ch_5	Glu–Ser–Lys–Tyr	4
	129 140	
Ch_{4-5}	Leu–Cys–Asn–Val–Ser–Glu–Val(Leu, Glu, Ser, Lys) Tyr	
	141	
Ch_6	Arg	1
		28

Determination by Edman degradation [4] $CN_{II}\,CT_{III}$: C-terminal tryptic fragment of citraconylated CN_{II} T : tryptic peptides of uncitraconylated $CN_{II}\,CT_{III}$; CH : chymotryptic peptides of $CN_{II}\,CT_{III}$.

Table 3
The amino acid sequence of viper hemoglobin α-chain



(Residues which are invariant in mammalian, chichen, carp and viper α-chains are in italics; helices A, B, etc . . . are indicated as determined for horse α-chain [11])

a C-terminal lysine, T_{14} has 14 residues and a C-terminal arginine. T_{13} is sequenced by Edman degradation and the first 7 residues of T_{14} are determined in the same way (table 1). A chymotryptic hydrolysis of the peptide CN_{II} CT_{II} gives 6 peptides which are purified by peptide mapping and partially sequenced. The results, shown in table 1, allow to establish the complete sequence of peptide CN_{II} CT_{II}.

2.3. *Determination of the amino acid sequence of peptide CN_{II} CT_{III}*

The C-terminal fragment of α-chain, CN_{II} CT_{III}, comprises 28 residues, two of which are lysine and one arginine in C-terminal position. After removal of citraconyl groups, CN_{II} CT_{III} (11 mg, about 4 μmoles) is split by trypsin into three peptides which are purified by chromatoelectrophoresis and analyzed. These three peptides correspond to the tryptic units T_{15}, T_{16} and T_{17} obtained from the intact α-chain. Peptide T_{15}, which contains 14 residues, has the same N-terminal sequence as intact CN_{II} CT_{III}; T_{16} comprises 12 residues one of which is lysine; T_{17} has 2 residues, tyrosine and arginine and therefore is C-terminal.

Edman degradation is applied on CN_{II} CT_{III} and on T_{15}, T_{16} and T_{17}. A chymotryptic hydrolysis of CN_{II} CT_{III} gives 8 fragments which are isolated, analyzed and partially sequenced. From the results the complete amino acid sequence of the fragment CN_{II} CT_{III} can be deduced (table 2).

Table 3 shows the sequence of the C-terminal part of α-chain determined in the present work and the complete sequence of viper α-chain as well.

3. Discussion

Because viper α-chain has a single methionine in position 32, cleavage with cyanogen bromide gives, on one hand a 32-residue N-terminal peptide, and on the other hand a large C-terminal fragment.

Sequence analysis of the N-terminal peptide is relatively easy because it contains 2 lysines and 3 arginines and cleavage by trypsin yields rather small peptides. The large C-terminal fragment of 109 residues has 10 lysines and 3 arginines one of which is in C-terminal position: therefore tryptic cleavage of the citraconylated fragment gives three peptides containing 60, 21 and 28 residues, respectively, which are purified by molecular sieving. The positions of these three peptides in the large C-terminal fragment can be deduced from their N- and C-terminal sequences. After removal of citraconyl groups and tryptic hydrolysis, they are split into 8, 2 and 3 peptides, respectively, which are isolated by chromatoelectrophoresis and partially sequenced by Edman degradation. Chymotryptic hydrolysis gives overlapping peptides which are purified and sequenced in the same way. From the results the complete sequence of the viper α-chain can be deduced.

When viper α-chain is compared to mammalian, chicken and carp α-chains [9], it can be noted that 46 residues out of 141 (about one third) are invariant in the 18 species so far investigated (table 3). Two residues, regarded as invariant before this work [9], are substituted in viper α-chain: lysine-11 (A_9) is replaced by arginine and histidine-122 (H_5) by leucine. The first substitution is conservative but not the second; because His-122 is supposed to play a role in the Bohr effect [10], it would be of interest to check whether viper hemoglobin displays this effect.

According to Perutz et al. [11], residues involved in the contacts either with heme or with β_2-chain are 'vital' for the biological function and therefore should be invariant. 10 residues of α-chain are involved in the contacts $\alpha_1\ \beta_2$. In viper they are identical with those found in human α-chain except valine-96 (G_3) which is replaced by alanine as in carp α-chain [12]. 8 are invariant in all the species investigated. The 19 residues of α-chain involved in the contacts with heme are identical in viper with those found in human α-chain but only 14 are invariant in all the species. 'Vital' contacts are located in the middle and C-terminal parts of the sequence [11]. When compared to human α-chain, 50% of substitution are found in the N-terminal part (32 residues) 32% in the middle (60 residues) and 31% in the C-terminal (49 residues) parts. The overall percentages of substitutions in chicken, viper, and carp α-chains are 25, 35 and 50, respectively, when human α-chain is taken as the reference; the approximate times of divergence of their respective ancestors, assessed from the fossil evidence [13], are 280, 300 and 450 million years.

Acknowledgement

The authors wish to thank Miss Christiane Devaux and Miss Monique Bourdin for their skilled technical assistance. This investigation was supported in part by grants from C.N.R.S. and D.G.R.S.T.

References

[1] Duguet, M., Chauvet, J. P. and Acher, R. (1973) FEBS Letters 29, 10–14.
[2] Duguet, M., Chauvet, J. P. and Acher, R. (1974) FEBS Letters 44, 83–86.
[3] Gross, E. and Witkop, B. (1962) J. Biol. Chem. 237, 1856–1860.
[4] Chauvet, J. P. and Acher, R. (1972) Biochemistry 11, 916–927.
[5] Spackman, D. H., Stein, W. H. and Moore, S. (1958) Anal. Chem. 30, 1190–1206.
[6] Duguet, M., Chauvet, J. P. and Acher, R. (1971) FEBS Letters 18, 185–188.
[7] Dixon, H. B. F. and Perham, R. N. (1968) Biochem. J. 109, 312–313.
[8] Acher, R. and Crocker, C. (1952) Biochim. Biophys. Acta 9, 704–705.
[9] Dayhoff, M. O. (1972) Atlas of Protein Sequence and Structure, Vol. 5, The National Biomedical Research Foundation, Silver Spring, Md.
[10] Perutz, M. F. (1970) Nature 228, 726–739.
[11] Perutz, M. F., Muirhead, H., Cox, J. M. and Goaman, L. C. G. (1968) Nature 219, 131–139.
[12] Hilse, K. and Braunitzer, G. (1968) Z. Physiol. Chemie 349, 433–450.
[13] Romer, A. S. (1966) Vertebrate Paleontology, The University of Chicago Press, Chicago, III.