THE β-CHAIN OF FROG HEMOGLOBIN (*RANA ESCULENTA*): THE COMPLETE AMINO ACID SEQUENCE

Jean-Pierre CHAUVET and Roger ACHER
Laboratory of Biological Chemistry, Faculty of Science, Paris, France

Received 6 August 1970

In two previous notes, the amino acid sequence of the first 98 residues of the β-chain of frog hemoglobin (*Rana esculenta*) has been described [1, 2]. This paper gives the sequence of the *C*-terminal part of the chain.

The carboxymethylated [3] and trifluoroacetylated [4] β-chain has been subjected to tryptic hydrolysis (enzyme/substrate weight ratio 1/100, 0.1 M ammonium bicarbonate pH 8.0, 3 hr, 37°). The five fragments TF_I, TF_{II}, TF_{III}, TF_{IV} and TF_V produced by cleavage at the 4 arginine residues were purified by gel-filtration on Sephadex G-50, then by paper electrophoresis or another gel-filtration on Sephadex G-25. The peptides TF_I, TF_{II} and TF_{III} account for the first 98 residues of the chain [1, 2]. The amino acid compositions [5] of TF_{IV} and TF_V account for the next 42 residues.

The peptide TF_{IV} contains 12 residues a single of which is basic, namely an arginine residue. Because the *C*-terminal sequence of the chain is Ala-Tyr-His [6], this peptide cannot be the *C*-terminal trifluoroacetylated fragment. TF_{IV} is identical to a peptide which can be isolated from the tryptic digest of the non-trifluoroacetylated β-chain. This tryptic unit was called T_{12} because of its position in the chain.

TF_V contains 30 residues one of which is lysine. After removal of the trifluoroacetyl group [4], this fragment is split by trypsin and gives two peptides T_{13} and T_{14}, containing respectively 27 and 3 residues, which can be recognized in the tryptic digest of the non-fluoroacetylated β-chain. T_{13} and T_{14} are separated by paper chromatoelectrophoresis [7]. Because the tripeptide T_{14} has the sequence Ala-Tyr-His, TF_V is obviously the *C*-terminal fragment of the β-chain. The amino acid sequence of T_{12}, T_{13} and T_{14} was determined by using Edman degradation [8] either directly on the fragments or on peptides obtained by chymotryptic hydrolysis (enzyme/substrate weight ratio 2/100, 0.1 M ammonium bicarbonate pH 8.0, 3 hr, 37°). Table 1 gives the results.

The positions of the tryptic units T_{12}, T_{13} and T_{14} in the chain can be confirmed by isolating overlapping fragments from the chymotryptic digest of the β-chain; the peptide Arg-Leu-Leu-Gly-Asn-Val-Phe gives the alignment T_{11}–T_{12}, the peptide Ala-Arg-His-Phe the alignment T_{12}–T_{13} and the peptide Ala-Lys-Ala-Tyr the alignment T_{13}–T_{14}.

The complete sequence of the β-chain of frog hemoglobin is shown in table 2. The chain comprises 140 residues. When it is compared to the human β-chain, six residues are deleted in the *N*-terminal portion of the frog β-chain. There are 61 substitutions so that 67 positions out of 146 of the β-chain pattern are modified. This number is greater than that found for mammalian β-chain (between 1 and 34 modifications) [9]. 70 positions are constant in twenty four mammalian β, δ and γ chains [9] but this number is reduced to 54 when the frog β-chain, which is so far the only non-mammalian β-chain fully known, is included in the comparison. It is noteworthy that there are two nearly immutable long sequences between the positions 24 and 32 (30 and 38 in the β-chain pattern) on the one hand, and between the positions 90 and 102 (96 and 108) on the other). This feature is to put together with the constant sequence between residues 70 and 80 in the cytochromes *c* of twenty five species [10].

North-Holland Publishing Company – Amsterdam

Table 1
Determination of the *C*-terminal sequence of the β-chain.

Peptide	Sequence	Number of residues
TF_{IV} T_{12}	99 110 Leu–Leu–Gly–Asn–Val–Phe–Ile–Thr–Val–Leu–Ala–Arg $\leftarrow T_{12}Ch_1 \rightarrow \leftarrow T_{12}Ch_2 \rightarrow \leftarrow T_{12}Ch_3 \rightarrow$	12
T_{13}	111 His–Phe–Gln–His–Glu–Phe–Thr–Pro–Glu–Leu–Gln–His–Ala–Leu– $\leftarrow T_{13}Ch_1 \rightarrow \leftarrow T_{13}Ch_2 \rightarrow$	27
TF_V	137 Glu–Ala–His–Phe–Cys–Ala–Val–Gly–Asp–Ala–Leu–Ala–Lys $\leftarrow T_{13}Ch_3 \rightarrow \leftarrow T_{13}Ch_4 \rightarrow$	
T_{14}	138 140 Ala–Tyr–His	3

→ Determination by Edman degradation: arrow above the sequence degradation performed directly on tryptic units; arrow under the sequence degradation performed on chymotryptic fragments of tryptic units. $T_{12}Ch_1$: chymotryptic fragment of peptide T_{12} etc....

Table 2
The amino acid sequence of the frog β-chain. (Residues which are constant in the mammalian β, δ, γ chains, presently known and the frog β-chain are in italics).

1 10 20
Gly–Ser–Asp–Leu–Val–Ser–Gly–Phe–Trp–Gly–*Lys*–*Val*–Asp–Ala–His–Lys–Ile–*Gly*–Gly–*Glu*

21 30 40
Ala–*Leu*–Ala–*Arg*–*Leu*–*Leu*–Val–*Val*–*Tyr*–*Pro*–*Trp*–*Thr*–Gln–*Arg*–Tyr–*Phe*–Thr–Thr–*Phe*–*Gly*

41 50 60
Asn–*Leu*–Gly–Ser–Ala–Asp–*Ala*–Ile–Cys–His–*Asn*–Ala–*Lys*–*Val*–Leu–Ala–*His*–*Gly*–Glu–*Lys*

61 70 80
Val–*Leu*–Ala–Ala–Ile–Gly–Glu–Gly–Leu–Lys–His–Pro–Glu–Asn–*Leu*–*Lys*–Ala–His–Tyr–Ala

81 90 100
Lys–*Leu*–*Ser*–*Glu*–Tyr–*His*–Ser–Asn–Lys–*Leu*–*His*–*Val*–*Asp*–*Pro*–Ala–*Asn*–*Phe*–Arg–*Leu*–*Leu*

101 110 120
Gly–*Asn*–Val–Phe–Ile–Thr–*Val*–*Leu*–Ala–Arg–His–Phe–Gln–His–Glu–*Phe*–Thr–*Pro*–Glu–Leu

121 130 140
Gln–His–Ala–Leu–Glu–Ala–His–Phe–Cys–Ala–*Val*–Gly–Asp–*Ala*–*Leu*–Ala–Lys–Ala–*Tyr*–*His*

Acknowledgement

The authors wish to thank Miss Christiane Devaux for her skilled technical assistance. This investigation was supported in part by a grant from the Délégation à la Recherche Scientifique et Technique (no. 6600146).

References

[1] J.P.Chauvet and R.Acher, FEBS Letters 8 (1970) 163.
[2] J.P.Chauvet and R.Acher, FEBS Letters 9 (1970) 202.
[3] A.M.Crestfield, S.Moore and W.H.Stein, J. Biol. Chem. 238 (1963) 622.
[4] R.F.Goldberger and C.B.Anfinsen, Biochemistry 1 (1962) 401.
[5] D.H.Spackman, W.H.Stein and S.Moore, Anal. Chem. 30 (1958) 1190.
[6] J.P.Chauvet and R.Acher, FEBS Letters 5 (1968) 305.
[7] C.Baglioni, Biochim. Biophys. Acta 48 (1961) 489.
[8] H.Fraenkel-Conrat, J. Am. Chem. Soc. 76 (1954) 3606.
[9] Atlas of Protein Sequence and Structure, ed. M.O. Dayhoff (1969) The National Biomedical Research Foundation.
[10] E.L.Smith, The Harvey Lectures 62 (1968) 231.