

BBA 36995

THE PRIMARY SEQUENCE OF CHICKEN MYOGLOBIN (*Gallus gallus*)

MARC DECONINCK^a, SERGE PEIFFER^b, JOSIANE DEPRETER, CLAUDINE PAUL,
ARTHUR GEORGES SCHNEK and JOSÉ LEONIS

*Laboratoire de Chimie Générale I, Faculté des Sciences, Université libre de Bruxelles, B-1050 Bruxelles
(Belgium)*

(Received October 23rd, 1974)

SUMMARY

After enzymatic digestion of chicken myoglobin by trypsin, chymotrypsin or thermolysin, the separation of peptides was performed by column chromatography on various ion exchange resins.

Each peptide was purified by high-voltage paper electrophoresis or by chromatography either on paper or on ion-exchange resin, and its complete amino acid sequence was then determined by the combined dansyl-Edman procedure and by endopeptidase digestions.

The whole globin was submitted to automatic Edman degradation using the Beckman sequencer.

Residues have been positioned from overlaps of sequence data between tryptic (T), chymotryptic (C) and thermolysin (Th) peptides. The stepwise degradation of the whole globin confirmed the alignment of the N-terminal third of the molecule.

The combination of these different approaches has led to the complete determination of the 153 residues sequence forming the polypeptide chain of chicken myoglobin.

Comparison of the established chicken myoglobin structure with those from other species shows a conservation of structure, although the avian protein exhibits more variations in its amino acid sequence than has been found between other known myoglobins which all belong to mammalian species.

INTRODUCTION

Since Edmundson [1] determined the amino acid sequence of the sperm whale myoglobin in 1965, this protein has been examined in several other species (land and marine mammals, gastropod).

As the complete covalent structures of myoglobins now available were con-

^a Present address: Laboratoire de Biochimie, Faculté de Médecine, Université Nationale du Zaïre-Campus de Kinshasa BP 171 Kin XI Rép. du Zaïre.

^b To whom requests for reprints should be addressed at the Laboratoire de Chimie Générale I, Faculté des Sciences, Université libre de Bruxelles, 50, av. F. D. Roosevelt, B-1050 Bruxelles, Belgique.

fined to proteins from mammalian species either from land mammals (horse [2], cattle [3], kangaroo [4], human [5], sheep [6], gibbon [7], baboon [8], macaco [8], and badger [9]) or from aquatic mammals (sperm whale [1], dolphin [10], seal and porpoise [11]), our intention was to extend our knowledge of myoglobins to samples belonging to another zoological class: the birds.

This study has been undertaken with a double purpose; firstly to make comparisons at the amino acid sequence level and if possible at the three dimensional level with mammalian myoglobins, and secondly, to try to make progress in the elucidation of the evolution of the avian class.

EXPERIMENTAL

Material

Chicken myoglobin was prepared from homogenized heart suspension by ammonium-sulfate fractionation; the main component was purified by gel filtration and chromatography on CM-Sephadex as previously described [12]. After removal of the heme moiety from the protein, the globin, denatured by heat for 10 min at 90 °C, was submitted to enzymatic hydrolysis.

Chymotryptic, tryptic and thermolysin digestions

Tryptic and chymotryptic digestions were performed as previously described [13]. Tryptic hydrolysis was terminated by lowering the pH to 2.1 with azetropic HCl. After centrifugation, an insoluble fraction was isolated.

Thermolysin digestion of the globin (200 mg) was carried out in a pH-stat at 40 °C and pH 8.0 with an enzyme/substrate weight ratio of 1:100. The reaction was stopped after 90 min by the addition of glacial acetic acid to pH 2.7. No significant insoluble material was detected after thermolysin and chymotryptic digestions.

Separation of peptides

Chymotryptic and tryptic peptides were separated by column chromatography on Dowex AG 50 X8 resin as described by Bradshaw et al. [14].

The insoluble tryptic fraction was fractionated on Dowex 1 X2 resin using experimental procedures related by the same authors [14].

The separation of thermolysin peptides was performed by column chromatography (25 × 2.0 cm) on Aminex A5 (Bio Rad) resin, at 40 °C. The column was developed successively with: (a) 250 ml of pyridine acetate 0.02 M pH 2.3; (b) a double linear gradient (1 l) of pyridine acetate 0.02 M pH 2.3 to 0.5 M, pH 3.9; (c) a double linear gradient (1 l) of pyridine acetate 0.5 M, pH 3.9 to 2.0 M, pH 5.0; (d) 250 ml of pyridine acetate 2.0 M, pH 5.0; (e) 300 ml of pyridine acetate 5.0 M, pQ 5.5.

A flow rate of 40 ml/h was maintained and the eluant was collected in 5-ml fractions.

All peptide fractions were pooled, concentrated by evaporation under vacuum or lyophilization and redissolved in water or in 1 % acetic acid.

Purification of peptides

Each fraction was purified by high voltage paper electrophoresis (Whatman

3MM) at pH 3.6 or 6.5 [15] and/or by paper chromatography using solvents reported by Light and Smith [16]. In some case, purification was performed by column chromatography (0.9×15.0 cm) at 50 °C on Aminex A5 resin using a double linear gradient of pyridine acetate (400 ml; pH 3.1 to 5.0 and 0.2 to 2.0 M in pyridine).

Sequence analysis

Amino acid analysis of the peptides was performed on an automatic amino acid analyzer (Beckman Unichrom).

Peptide sequences were determined by the combined dansyl-Edman technique [17, 18]. Dansyl amino acids were identified by two-dimensional chromatography on polyamide sheets of 5.0×5.0 cm using solvents reported by Woods and Wang [19]. The C-terminal and N-terminal residues were identified by amino acid analysis after kinetic hydrolysis respectively with carboxypeptidases A and/or B and leucine aminopeptidase according to experimental procedures described by Han [20].

The whole globin was submitted to automatic Edman degradation [21] using the Beckman sequencer (model 890).

The identification of the phenylthiohydantoin (PTH) amino acids was realized by gas chromatography [22] (Beckman CG 45 chromatograph) and in some cases by automatic analysis of the amino acid liberated by acid hydrolysis (6.0 M HCl, 120 °C; 16 h) of the PTH derivatives [23].

Assignment of amide side chains

The identification of asparagine and glutamine was made from three types of experimental data: (a) electrophoretic mobility of the peptides at pH 6.5; (b) analysis of peptide hydrolyzates by carboxypeptidases A, B and leucine aminopeptidase; (c) gas chromatography of the PTH derivatives obtained from the sequencer degradation.

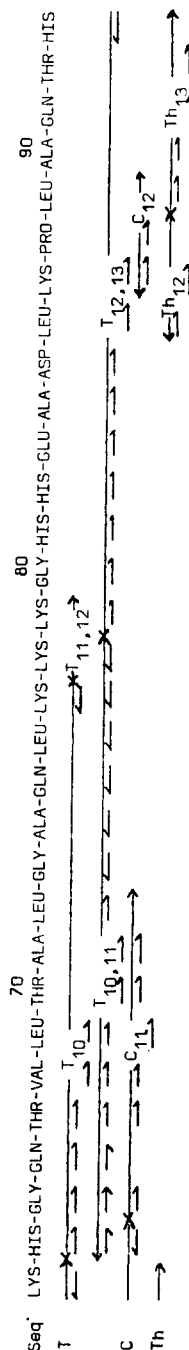
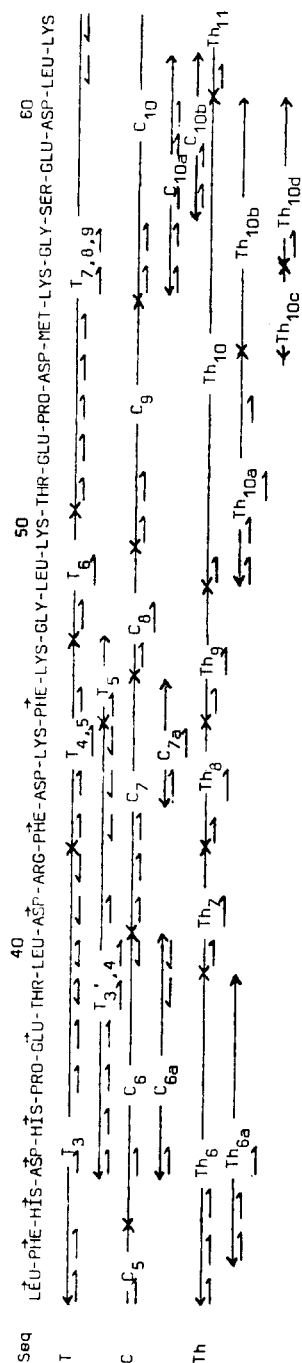
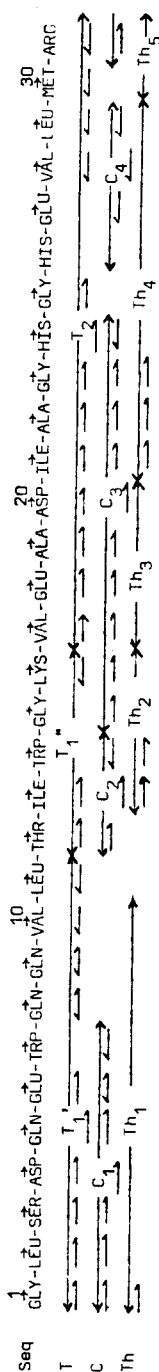
RESULTS

The amino acid composition of the tryptic, chymotryptic and thermolysin peptides are deposited with the BBA Data Bank*. In front of the inventory of the various amino acid sequences of the peptides** one can say that: (a) to the exclusion of two residues of lysine, the tryptic peptides account for the total number of arginine and lysine obtained by automatic amino acid analysis of the whole globin; (b) only one or two tryptic peptides corresponding to six residues in 153 have not been isolated; (c) each set of chymotryptic or thermolysin peptides accounts for about 70–80% of the total sequence; (d) the amino acid sequence of a chymotryptic peptide (C_{14}) does not correspond with any of the tryptic ones, which helped fix its position in the space not covered by the tryptic peptides.

From these results and taking the primary structure of sperm whale myo-

* Supplementary data to this article, giving details of amino acid composition of these peptides are deposited with and can be obtained from: —Elsevier Publishing Company, BBA Data Deposition, P.O. Box 1527, Amsterdam, The Netherlands. Reference should be made to No. BBA/DD/020/36995/386 (1975) 567.

** Any detail on the sequence determination appeared in Dr M. Deconinck's thesis presented at the University of Brussels and is available on request.



globin as a model, an alignment of the tryptic peptides has been made. The overlap of the chymotryptic and thermolysin peptides allows us to deduce without any ambiguity the complete amino acid sequence of chicken myoglobin which is shown in Fig. 1.

In order to facilitate presentation of the data, peptides are numbered in the order in which they appear in the sequence starting from the amino terminal end. The tryptic peptides (T) are counted after the number of the basic residues (Arg. and Lys). The notation T' indicates a peptide resulting from an unexpected split by trypsin. In the chymotryptic (C) and thermolysin (Th) peptides sets, a letter is added when a short peptide can be related to a longer one. A few particular points might be outlined:

Residues 1 to 47

The amino acid sequence of this N-terminal segment of chicken myoglobin up to residue 47 was almost completely established by characterization of the tryptic peptides. Their positions and sequences were confirmed by overlaps with chymotryptic and thermolysin peptides and by automatic sequential degradation of the whole globin.

No distinct tryptic peptide was found covering the residues 1 to 16. This peptide represents an insoluble core peptide of the protein. However as the leucyl bond at position 11 is split during tryptic digestion, two peptides covering these positions are identified.

Residues 48 to 63

Overlaps with various chymotryptic and thermolysin peptides confirm that the tryptic peptides T₆ and T_{7,8,9} can be assigned to these residues.

Residues 64 to 96

These residues are covered by tryptic peptides T_{10,11} and T_{12,13}. Peptide T_{10,11} supplies only a single residue overlap with peptide C₁₀ (a His at position 64). However these tryptic peptides are assigned on the basis of similar myoglobin sequences and although no rigorous overlap was obtained in this region the choice we have made is the only one permitted consistent with the proposed structure.

On the basis of identical N- and C-terminal structures (His . . . Lys-Lys) added to a same number of residues, the permutation of peptides T_{10,11} and T_{18,19} should be considered. An arrangement different from that shown is incompatible with the C-terminal part of the molecule established without any ambiguity.

Residues 97 to 113

Peptide Th₁₄ overlaps peptides T_{12,13} and C₁₄, thus establishing the sequence from residues 97 to 103. Peptide T₁₆, isolated from the insoluble tryptic fraction, covers the positions 103 to 113 supplying only a single residue overlap with peptide C₁₄ (Tyr at position 103). Such an overlap is consistent with the proposed structure because (a) the amino acid composition of the whole globin reveals the presence of only two residues of tyrosine, and (b) the peptide T_{22,23} which is the other peptide containing a residue of Tyr is fixed without any ambiguity at the C-terminal part of

the molecule on the basis of the identity of its C-terminal sequence with that obtained by hydrolysis of the whole globin by the carboxypeptidases A and B (... Gln-Gly).

Residues 114 to 153

Peptides T₁₇ and T_{18,19} overlapped by peptides C₁₆, C₁₈, Th₁₆ and Th₁₉ fix the positions from residues 114 to 135. No overlap at the N-terminal part of peptide T₁₇ was found. Peptide T₂₀, T₂₁ and T_{22,23} covered by various chymotryptic and thermolysin peptides established the C-terminal sequence of the molecule.

DISCUSSION

The peptide bonds were generally hydrolyzed according to the accepted specificity of the different proteases used, although certain unusual points of cleavage were observed. For example a cleavage after Leu-11 during digestion with trypsin was observed despite the fact that trypsin was previously treated with diphenyl carbamyl chloride.

The complete amino acid sequences of chicken and sperm whale myoglobins and the partial one of penguin [24] up to residue 70 are aligned in Fig. 2.

The chicken protein differs from the sperm whale myoglobin at 43 sites and from penguin myoglobin at 14 sites within the first 70 residues.

When compared with sperm whale globin, it appears that more than half of the substitutions observed are consecutive to conservative changes between chemically similar residues. Except for Thr/His at position 12, Gly/His at position 48, Ala/His at position 116, Ser/Ala at positions 127 and 144, other substitutions occur between the same types of amino acids, either between polar residues or between hydrophobic ones. When compared with all other complete sequences already determined (sperm whale, horse, seal, porpoise, cattle, kangaroo, dolphin, human, gibbon, baboon, macaco and badger), chicken myoglobin presents substitutions at six positions where no mutation has been observed before. These replacements are Gln for Gly at position 5, Met for Ile at position 30, Asp for Glu at position 41, Gly for His at position 48, Gln for Ile at position 75 and Ser for Ala at position 127.

The different residues observed when comparing chicken and sperm whale proteins, can be considered as having appeared most frequently as the result of a single point mutation [25]. The respective replacements of Gly-5, His-12, Lys-34, Ser-35, Val-66, Ile-75, His-113, His-116, Ser-117, and Lys-140 in sperm whale by Gln, Thr, His, Asp, Gln, Gln, Lys, Ala, Glu, Asp in chicken are dependent on the exchange of two bases in the codon triplet.

On the basis of the total number of amino acid differences between chicken globin and all myoglobins of known amino acid sequence, it appears that chicken haemoprotein is markedly different from mammalian myoglobins. However, as chicken myoglobin has about three times as much in common with mammalian myoglobins than with the α [26] or β [27] chains of chicken haemoglobin, it is much closer to myoglobins from other species than to avian haemoglobin.

An examination of the amino acid sequence of chicken myoglobin in relation to the three-dimensional structure of sperm whale shows that the replacements occur both at external and internal positions in almost any part of the molecule. The

differences observed inside are generally interchanges between very similar residues, essentially hydrophobic amino acids.

The primary structure reported in this paper thus extends the comparative studies of myoglobins to another zoological class: the birds. As sequence studies undertaken in parallel on penguin (*Aptenodytes forsteri*) myoglobin are in progress, we hope to be able in the very near future to make broader phylogenetic comparisons of avian myoglobins at the amino acid sequence level.

Support from the Fonds National de la Recherche Scientifique and from the Fonds de la Recherche Fondamentale Collective is gratefully acknowledged*. The

* This work has also been supported by a grant from the National Science Foundation (GB 18483).

authors are greatly indebted to Professor F. W. Putnam of the Indiana University for the loan of a sequencer, to Professor M. Dautrevaux and Dr K. K. Han from the Université de Lille for the helpful discussions and encouragements and to S. Popescu for valuable technical assistance.

REFERENCES

- 1 Edmundson, A. B. (1965) *Nature* 205, 883-887
- 2 Dautrevaux, M., Boulanger, Y., Han, K. and Biserte, G. (1969) *Eur. J. Biochem.* 11, 267-277
- 3 Han, K., Dautrevaux, M., Chaila, X. and Biserte, G. (1970) *Eur. J. Biochem.* 16, 465-471
- 4 Air, G. M. and Thompson, E. O. P. (1971) *Aust. J. Biol. Sci.* 24, 75-95
- 5 Romero-Herrera, A. E. and Lehmann, H. (1971) *Nat. New Biol.* 232, 149-152
- 6 Han, K., Tetaert, D., Moschetto, Y., Dautrevaux, M. and Kopeyan, C. (1972) *Eur. J. Biochem.* 27, 585-592
- 7 Romero-Herrera, A. E. and Lehmann, H. (1971) *Biochim. Biophys. Acta* 251, 482-488
- 8 Romero-Herrera, A. E. and Lehmann, H. (1972) *Biochim. Biophys. Acta* 278, 62-67
- 9 Tetaert, D., Han, K., Plancot, M., Dautrevaux, M., Ducastaing, S., Hombrados, I. and Neuzil, E. (1974) *Biochim. Biophys. Acta* 351, 317-324
- 10 Karadjova, M., Nedkov, P., Bakardjieva, A. and Genov, N. (1970) *Biochim. Biophys. Acta* 221, 136-139
- 11 Bradshaw, R. A. and Gurd, F. R. N. (1969) *J. Biol. Chem.* 244, 2167-2181
- 12 Deconinck, M., Peiffer, S., Schnek, A. G. and Léonis, J. (1972) *Biochimie* 54, 969-972
- 13 Deconinck, M., Depreter, J., Paul, C., Peiffer, S., Schnek, A. G., Putnam, F. W. and Léonis, J. (1972) *FEBS Lett.* 23, 279-281
- 14 Bradshaw, R. A., Garner, W. H. and Gurd, F. R. N. (1969) *J. Biol. Chem.* 244, 2149-2158
- 15 Katz, A. M., Dreyer, W. S. and Anfinsen, C. B. (1959) *J. Biol. Chem.* 234, 2897-2900
- 16 Light, A. L. and Smith, E. L. (1962) *J. Biol. Chem.* 237, 2537-2546
- 17 Dopheide, T. A. A., Moore, S. and Stein, W. (1967) *J. Biol. Chem.* 242, 1833-1837
- 18 Gray, W. R. and Hartley, B. S. (1963) *Biochem. J.* 89, 59
- 19 Woods, K. R. and Wang, K. T. (1967) *Biochim. Biophys. Acta* 133, 369-370
- 20 Han, K. K. (1970) Thèse de doctorat, Université de Lille
- 21 Edman, P. and Begg, C. (1967) *Eur. J. Biochem.* 1, 80-91
- 22 Pisano, J. J. and Bronzert, T. J. (1969) *J. Biol. Chem.* 244, 5597-5607
- 23 Van Orden, H. and Carpenter, F. (1964) *Biochem. Biophys. Res. Commun.* 14, 399-403
- 24 Peiffer, S., Deconinck, M., Paul, C., Depreter, J., Schnek, A. G. and Léonis, J. (1973) *FEBS Lett.* 37, 295-297
- 25 Nirenberg, M. W., Leder, P., Bernfield, M., Brimacombe, R., Trupin, J., Rottman, F. and O'Neal, C. (1965) *Proc. Natl. Acad. Sci. U.S.A.* 53, 1161-1168
- 26 Matsuda, G., Takei, H., Wu, K. C. and Shiozawa, T. (1971) *Int. J. Prot. Res.* 3, 173-174
- 27 Matsuda, G., Maita, T., Mizuno, K. and Ota, H. (1973) *Nat. New Biol.* 244, 244
- 28 Flory, J. P. (1969) in *Statistical Mechanics of Chain Molecules*, p. 270, Interscience, London