THE N-TERMINAL AND C-TERMINAL AMINO ACID SEQUENCE OF BADGER MYOGLOBIN

D. TETAERT, K. HAN and M. DAUTREVAUX
Laboratoire de Chimie Biologique, Faculté de Médecine, Place de Verdun, 59045, Lille Cedex, France

and

S. DUCASTAING, I. HOMBRADOS and E. NEUZIL
Biochimie médicale, U.E.R. III, Université de Bordeaux II, 146, Rue Léo Saignat, 33076, Bordeaux, France

Received 13 October 1972
Revised version received 10 November 1972

1. Introduction

A comparative study of the covalent structure of several animal myoglobins has been undertaken by some of us (Lille group): the complete covalent structure of horse [1], ox [2] and sheep [3] myoglobin has been determined and the partial amino acid sequence of hog [4] and dog [5] myoglobins investigated. No information has been reported on the myoglobin of the badger (*Meles meles*), a wild animal rather common in the Bordeaux region for which some biochemical data are already available [6, 7]. Therefore, in order to extend the structural studies and for the purpose of establishing evolutionary and genetic informations, we are attempting to deepen the comparative studies of this peculiar protein, using different data reported in the literature [8–16].

2. Experimental

The myoglobin was extracted from muscle according to the procedure previously reported [17] and submitted to tryptic hydrolysis by adding 4% (w/w) enzyme (trypsin, Seravac) to substrate solution (10 mg/ml) and allowing the enzymatic digestion to process at 38° for 150 min at pH 8.75. The cleavage of the globin with BrCN and the isolation of large peptides were performed under similar conditions as those previously reported for ox and sheep globins [2, 3]. The quantitative amino acid analysis of large and short peptides were performed in an automatic analyser (Jeolco, Type 5 AH). The tryptophan content of large peptides was determined by Spies and Chambers' method [18]; the tryptophan content of short peptides were hydrolysed in 5.6 N HCl containing 6% (v/v) concentrated thioglycollic acid [19]. The amino terminal residues of large peptides were identified by dansylation in the system described by Gros and Labouesse [20], whereas those of short peptides were determined by dansylation on polyamide sheets (5 × 5 cm) in the system reported by Hartley [21]. The complete sequence determination of some peptides appeared unnecessary when amino acid composition, N-terminal amino acid, paper electrochromatography migration and peptide elution pattern on resin proved identical with those of corresponding tryptic peptides of horse myoglobin. When the amino acid composition of a peptide was different from homologous horse tryptic peptide, its sequence was determined by dansyl–Edman technique [21, 22], hydrazinolysis, carboxypeptidases (A and B) and/or aminopeptidase M digestions.

The separation and isolation of short tryptic peptides deriving from the N-peptide and the C-peptide obtained by BrCN cleavage of the whole

North-Holland Publishing Company – Amsterdam

Table 1
Amino acid composition of N, M and C peptides of badger myoglobin.

	N peptide	M peptide	C peptide
Lys	6.83 (7)	10.28 (10)	3.70 (4)
His	2.24 (2)	3.24 (3)	
Arg	1.01 (1)		1.01 (1)
Asp	4.86 (5)	4.53 (4–5)	2.00 (2)
Thr	1.29 (1)	3.24 (3)	
Ser	2.20 (2)	4.16 (4)	
H Se*	0.37 (1)	0.39 (1)	
Glu	8.90 (9)	10.97 (11)	3.07 (3)
Pro	1.37 (1)	3.41 (3)	
Gly	6.44 (6–7)	7.78 (8)	2.42 (2)
Ala	2.83 (3)	9.59 (9–10)	3.05 (3)
Val	3.07 (3)	2.59 (3)	
He	1.01 (1)	3.23 (3)	1.00 (1)
Leu	7.33 (7)	8.22 (8)	2.93 (3)
Tyr		1.30 (1)	0.92 (1)
Phe	3.20 (3)	2.40 (2–3)	1.88 (2)
Trp	++		
Total number residues	55	76	22
N terminal amino acid	Gly	Lys	Lys

* H Se: homoserine; H Se results from the reaction of BrCN on methionyl bonds.

protein were achieved by column chromatography on resin chromobeads P(Technicon) (N-peptide) or by preparative fingerprint technique recently reported by us [23] (C-peptide).

3. Results

The badger myoglobin contains only 2 residues of methionine; the cleavage of methionine bonds by BrCN leads necessarily to 3 peptides: an N-termal peptide (N-peptide), a C-terminal peptide (C-peptide) and the medium segment (M-peptide).

The amino acid composition of these peptides is given in table 1.

The N-peptide and the C-peptide were submitted to tryptic digestion. The amino acid composition and N-terminal amino acid of each tryptic peptide is given in table 2.

Because of the close analogy between mammalian myoglobins, it was possible to make a tentative alignment of the isolated tryptic peptides with the horse protein sequence as a model.

The amino acid sequence of N-peptide (55 residues) and those of C-peptide (22 residues) are given in table 3.

4. Discussion

The complete sequence of N-terminal segments of badger myoglobin is shown in table 3 and the proposed structure is compared to that of horse myoglobin.

The amount of each amino acid in the N-peptide and in the C-peptide resulting from BrCN cleavage of whole protein is in good agreement with the amino acid established in sequence.

Some particular results might be interesting to be outlined and discussed:

i) We have not isolated the tryptic peptide located in the position from N-terminal glycine to lysine-16. This peptide represents the first tryptic insoluble core. The peptide N-T_1 (Gly–Leu–Ser–Asp–Gly–Glu–Trp–Gln–Leu–Val–Leu: position from 1 to 11 in the molecule) contains neither lysine nor arginine; therefore trypsin does split the leucyl bond at position 11. The peptide N-T_2 is Asn–Val–Trp–Gly–Lys (position: 12 to 15). Thus, the first tryptic insoluble core was cleaved by trypsin into 2 peptides which were soluble in a buffer solution. However, the yield of cleavage is only 25% if compared with normal tryptic peptides.

ii) The acid hydrolysis in the presence of 6% (v/v) concentrated thioglycollic acid for the tryptophan containing peptides (T-N_1 and T-N_2) yields very satisfactory results for the tryptophan preservation: 0.85 tryptophan recovery for N-T_1 peptide and 0.88 residue tryptophan recovery for N-T_2 peptide.

The acid hydrolysis in the presence of 2 drops of 5% phenol in aqueous solution preserves the tyrosine contained in the peptide.

We have obtained a relatively low recovery of homoserine after acid hydrolysis. The methionine residue is transformed to homoserine lactone during and after the cleavage of methionyl bonds by Br-CN [24].

Table 2
Amino acid compositions and N-terminal amino acid (dansyl) of tryptic peptides obtained from the N-peptide and C-peptide.

Peptides	N-T_1	N-T_2	N-T_3	N-T_4	N-T_5	N-T_6	N-T_7	N-T_7 bis	N-T_8	N-T_9	C-T_1	C-T_2	C-T_3	C-T_4	C-T_5
Asp	1.1 (1)	1.20 (1)	1.16 (1)			0.90 (1)		1.22 (1)		1.09 (1)			2.01 (2)		
Thr					0.93 (1)										
Ser	1.01 (1)									1.06 (1)					
Glu	2.11 (2)		3.00 (3)		2.00 (2)					2.29 (2)		1.06 (1)			2.00 (2)
Pro					1.00 (1)										
Gly	2.19 (2)	0.98 (1)	2.00 (2)		1.01 (1)										2.00 (2)
Ala			2.09 (2)									0.99 (1)	2.09 (2)		
Val	0.84 (1)	1.06 (1)	1.85 (2)												
Ile			0.88 (1)										0.94 (1)		
Leu	2.83 (3)		1.82 (2)	1.08 (1)	0.99 (1)				1.49 (1)			1.88 (2)			1.00 (1)
Tyr														1.20 (1)	
Phe				1.13 (1)		1.00 (1)	1.04 (1)	1.86 (2)				0.95 (1)			0.99 (1)
Lys		0.89 (1)		0.80 (1)	1.11 (1)	1.12 (1)	0.96 (1)	1.91 (2)	0.81 (1)		2.00 (2)		0.97 (1)	0.80 (1)	
His			1.26 (1)		0.84 (1)				0.71 (1)						
Arg			0.94 (1)									1.12 (1)			
Trp	0.85 (1)	0.88 (1)													
Hse										0.55 (1)					

Table 2 (continued)

	$N\text{-}T_1$	$N\text{-}T_2$	$N\text{-}T_3$	$N\text{-}T_4$	$N\text{-}T_5$	$N\text{-}T_6$	$N\text{-}T_7$	$N\text{-}T_7$ bis	$N\text{-}T_8$	$N\text{-}T_9$	$C\text{-}T_1$	$C\text{-}T_2$	$C\text{-}T_3$	$C\text{-}T_4$	$C\text{-}T_5$
Total residues of amino acids	11	5	15	3	8	3	2	5	3	5	2	6	6	2	6
Position	1–11	12–16	17–31	32–34	35–42	43–45	46–47	43–47	48–50	51–55	132–133	134–139	140–145	146–147	148–153
N-terminal amino acid	Gly	Asp	Val	Leu	Gly	Phe	Phe	Phe	His	Ser	Lys	Ala	Asp	Tyr	Glu

The composition of each peptide is given as the molar ratios of the amino acid. The numbers in the parentheses indicate the nearest integral numbers. The amino acid analyses of those peptides are given in the order in which the peptides occur in the molecule and have been numbered accordingly. The letter N designates the peptides occurring in the N-terminal segment, the letter C indicates the peptides occurring in the C-terminal segment in the myoglobin molecule. The letter T designates the peptides resulting from trypsin digestion.

iii) The lysyl bond situated at position 45 is partially cleaved by trypsin. Therefore we have obtained 3 peptides: $N\text{-}T_6$ (Phe–Asp–Lys), $N\text{-}T_7$ (Phe–Lys) and the overlapping peptide peptide $N\text{-}T_{7bis}$ (Phe–Asp–Lys–Phe–Lys) (position 43 to 47). The presence of an aspartic acid just before lysine residue (position 45) may cause the partial hydrolysis of this lysyl bond by trypsin.

In the N-terminal segment, the badger protein differs from horse protein at 5 sites and in the C-terminal segment only 1 difference was observed. Furthermore all the variations observed among the two proteins are confirmed to amino acid interchanges caused by changing only one base in the coding triplet [25].

The primary structure of the badger myoglobin is that of a typical "mammalian-type" myoglobin, namely, a single polypeptide chain 153 residues long, with a glycine amino-terminal residue, and the heme prosthetic group "bonded" to two histidine residues in position 64 and 93. No notable difference between the badger protein and the other mammalian myoglobin is found on the N-terminal and C-terminal segments because the substitution of all 6 residues involves one base change in the codon. The differences in sequence between the myoglobin of the hooved animals (horse, beef and sheep) and carnivore (badger) with which they have a presumed common origin, are found mainly in residues 9, 21, 34, 51, 53 and 132; of these the changes in the position 9 (Gln-horse to Leu-badger) is the most notable substitution, because leucine-9 in badger is also existing in the other mammalian myoglobins except in equine and bovine myoglobins.

The difference observed in positions 21 and 51 between horse and badger proteins are "conservative" mutations (Ile-21:horse to Leu-21:badger and Thr-51:horse to Ser-51:badger). It is also interesting to point out to "punctual" substitutions among the horse and badger myoglobins at positions 53 (Ala to Asp) and 132 (Thr to Lys).

A similar substitution at position 132 has been observed among mammalians (Ser in beef and sheep), Asn in sperm whale. However, it is difficult to evaluate phylogentically the distance from badger to hooved animals, to cetaceans and primates owing to the complete covalent structure of badger myoglobin is not yet established. We will consider in detail the phylogency and evolution of myoglobins in the next paper.

Table 3

Comparison of the amino acid sequence of N-terminal and C-terminal segments of horse and badger myoglobin. The non identical residue of badger are underlined.

N-peptide

```
          1                   5                  10                  15                  20
Horse : Gly-Leu-Ser-Asp-Gly-Glu-Trp-Gln-Gln-Val-Leu-Asn-Val-Trp-Gly-Lys-Val-Glu-Ala-Asp-
Badger: Gly-Leu-Ser-Asp-Gly-Glu-Trp-Gln-Leu-Val-Leu-Asp-Val-Trp-Gly-Lys-Val-Glu-Ala-Asp-

          21                  25                  30                  35                  40
Horse : Ile -Ala-Gly-His -Gly-Gln-Glu-Val-Leu-Ile-Arg-Leu-Phe-Thr-Gly-His-Pro-Glu-Thr-Leu-
Badger: Leu-Ala-Gly-His -Gly-Gln-Glu-Val-Leu-Ile-Arg-Leu-Phe-Lys-Gly-His-Pro-Glu-Thr-Leu-
          41                  45                  50                  55
Horse : Glu-Lys-Phe-Asp-Lys-Phe-Lys-His-Leu-Lys-Thr-Glu-Ala-Glu-Met (Hse).
Badger: Glu-Lys-Phe-Asp-Lys-Phe-Lys-His-Leu-Lys-Ser-Glu-Asp-Glu-Met (Hse).
```

C-peptide

```
          132                 135                 140                 145                 150
Horse : Thr-Lys-Ala-Leu-Glu-Leu-Phe-Arg-Asn-Asp-Ile -Ala-Ala-Lys-Tyr-Lys-Gly-Leu-Gly-Phe
Badger: Lys-Lys-Ala-Leu-Glu-Leu-Phe-Arg-Asn-Asp-Ile -Ala-Ala-Lys-Tyr-Lys-Gly-Leu-Gly-Phe
          152 153
Horse : Gln-Gly-COOH
Badger: Gln-Gly-COOH
```

In this paper we have determined the amino acid sequence of N-terminal and C-terminal peptides of badger myoglobin; 77 residues out of 153 were characterised by us.

Acknowledgement

Our thanks are due to Prof. Canivenc, Université de Bordeau II, for supplying animals.

References

[1] M. Dautrevaux, Y. Boulanger, K. Han, G. Biserte, European J. Biochem. 11 (1969) 267.
[2] K. Han, M. Dautrevaux, X. Chaila and G. Biserte, European J. Biochem. 16 (1970) 465.
[3] K. Han, D. Tetaert, Y. Moschetto, M. Dautrevaux and C. Kopeyan, European J. Biochem. 27 (1972) 585.
[4] V. Dumur, M. Dautrevaux and K. Han, FEBS Letters 26 (1972) 241.
[5] R. Floc'h, M. Dautrevaux and K. Han, Biochimie, submitted to publication.
[6] F. Tayeau, J. Marquevielle, R. Nivet and M. Dumas, Bull. Soc. Pharm. Bordeau 98 (1959) 106.
[7] S. Ducastaing and E. Neuzil, C.R. Soc. Biol. 163 (1969) 2140.
[8] A.B. Edmunson, Nature 205 (1965) 389.
[9] Bradshaw and R.R.N. Gurd, J. Biol. Chem. 244 (1969) 267.
[10] M. Karadjova, P. Nedkov, A. Bakardjieva and N. Genov, Biochem. Biophys. Acta 221 (1970) 136.
[11] A.E. Romero-Herrera and H. Lehmann, Nature New Biology 232 (1971) 149.
[12] C.M. Air, E.O.P. Thompson, B.J. Richardson and G.B. Sharman, Nature (London) 229 (1971) 391.
[13] A.E. Romero-Herrera and H. Lehmann, Biochem. Biophys. Acta 251 (1971) 482.
[14] M. Deconinck, J. Depreter, C. Paul, S. Pfeiffer, A.G. Schnek, F.W. Putnam and J. Leonis, FEBS Letters 23 (1972) 279.
[15] W. Votsch and F.A. Anderer, Z. Naturforschung 27b (1972) 157.
[16] A.E. Romero-Herrera and H. Lehmann, Biochem. Biophys. Acta 278 (1972) 62.
[17] M. Dautrevaux, V. Dumur and K. Han, Biochimie 53 (1971) 717.
[18] J.R. Spies and D.C. Chambers, Anal. Chem. 21 (1949) 106.
[19] H. Matsubara and R. Sasaki, Biochem. Biophys. Res. Commun. 35 (1969) 175.
[20] C. Gros and B. Labouesse, European J. Biochem. 7 (1969) 463.
[21] W.R. Gray and B.S. Hartley, Biochem. J. 89 (1963) 59.
[22] B.S. Hartley, Biochem. J. 119 (1970) 805.
[23] K. Han, B. Debuire, M. Dautrevaux, G. Biserte, A. Fattoum, F. Regnouf, R. Kassab and L.A. Pradel, Compt. Rend. 274 (1972) 324.
[24] E. Gross and B. Witkop, J. Biol. Chem. 237 (1962) 1856.
[25] M.W. Niremberg, P. Leder, M. Bernfeld, R. Brimacombe, J. Trupin, F. Rotteman and C.O'Neal, Proc. Natl. Acad. Sci. U.S. 53 (1965) 1161.