PRIMARY STRUCTURE OF APLYSIA MYOGLOBIN: SEQUENCE OF A 63-RESIDUE FRAGMENT

L.TENTORI, G.VIVALDI, S.CARTA, M.MARINUCCI and A.MASSA
Laboratori di Biologia, Instituto Superiore di Sanità, Roma, Italia

and

E.ANTONINI and M.BRUNORI
Institute of Biochemistry, University of Rome and "Regina Elene" Institute, Rome, Italy

Received 28 October 1970

The myoglobin contained in the buccal muscle of the Mediterranean Gasteropod *Aplysia limacina* shows a number of interesting functional and structural features when compared with mammalian myoglobins. The chemical and physico-chemical properties of the protein, equilibrium and kinetic constants for the reaction of this pigment with oxygen and carbon monoxide, and its amino acid composition and tryptic peptide pattern have been previously reported [1–4]. The present paper deals with the determination of the sequence of 63 residues at the carboxyl terminal end.

Aplysia myoglobin was prepared as described earlier [1]. Further purification was obtained by DEAE-cellulose chromatography which resolves the pigment into two heme-containing components; a minor one, corresponding to about 10% of the total, which was not used, and a major one (about 90%).

Digestion with carboxypeptidases showed that the globin of the purified major component has a carboxyl terminal sequence Ala–Ala–Gly–Lys; on the other hand attempts to detect the amino terminal residue were unsuccessful. On the assumption that acetylation occurs at the amino terminal of the amino acid, the acetic acid content of the myoglobin was quantitatively estimated by the gas-chromatographic technique of Ward et al. [5]. The analysis yielded 0.87 acetyl moles per mole of globin, thus proving the presence of an acetylated residue in the amino terminal position of the polypeptide chain.

The globin was subjected to CNBr cleavage, with a technique similar to that described by Gross and Witkop [6], in 70% formic acid for 24 hr at 28° using a methionine–CNBr ratio of 1:50. In keeping with the fact that Aplysia myoglobin contains three methionine residues, four fragments were obtained: M_1CNBr, M_2CNBr, M_3CNBr and M_4CNBr. The separation of CNBr fragments was by gelfiltration with 5 × 200 cm Sephadex G-75 columns eluted with 0.2 M acetic acid.

Table 1
Amino acid composition of Aplysia myoglobin cyanogen bromide fragments.

	M_1CNBr	M_2CNBr	M_3CNBr	M_4CNBr
Lys	9	3	1	–
His	–	–	1	–
Arg	3	–	1	–
Asp	15	2	–	–
Asn	–	–	1	–
Thr	1	1	–	–
Ser	8	1	3	1
Glu	3	–	2	–
Gln	–	–	2	–
Pro	3	3	–	–
Gly	4	4	3	–
Ala	18	9	2	1
Val	6	2	3	–
ILe	2	2	–	–
Leu	8	3	1	–
Phe	11	3	3	–
HSR	1	–	1	1
Trp	+	1	–	–
Total	91	34	24	3

Table 2
Amino terminal and carboxyl terminal sequences of Aplysia myoglobin cyanogen bromide fragments.

	Amino terminal sequences	Carboxyl terminal sequences
M_1CNBr	Acetyl x	←Lys–HSR
M_2CNBr	→Phe–Pro–Gly–Phe–Val–Ala–Ser–Val	←Ala–Ala–Gly–Lys
M_3CNBr	→Leu–Ser	←Val–Arg–Ser–HSR
M_4CNBr	→Ser–Ala–HSR	

→ determination by dansyl-Edman degradation; ← determination by carboxypeptidase digestion

The results of the quantitative amino acid analyses performed on the isolated fragments with the automatic method of Spackman et al. [7], are shown in table 1. The carboxyl terminal and amino terminal sequences of each fragment were determined by carboxypeptidase A (CPA) plus carboxypeptidase B (CPB) digestion [8] and dansyl-Edman degradation [9–11], respectively; the results are shown in table 2.

Dansyl-Edman degradation of M_4CNBr revealed its complete sequence: Ser–Ala–HSR (HSR = homoserine). M_1CNBr was identified as the amino terminal fragment of the chain because: (i) attempts to detect its alpha-amino group were unsuccessful; (ii) gas-chromatographic analysis according to Ward et al. [5] yielded one mole of acetic acid per mole. On the other hand M_2CNBr was identified as the carboxyl terminal end of the polypeptide chain because its carboxyl terminal sequence is identical to that of whole globin, and homoserine is absent. The overall order of the fragments was finally determined by submitting the whole globin to tryptic digestion. The resulting tryptic peptides were separated with the chromatographic technique of Jones [12], and, when necessary, purified by rechromatography on Dowex 50 × 2, 0 × 150 cm columns operated with a linear gradient obtained by means of a 0.2 M pyridine–acetic acid starting buffer, pH 3.1, and a 2.0 M pyridine–acetic acid limiting buffer, pH 5.0. The methionine-containing peptides were detected with the aid of sulphur reaction on paper strips [13]. (Aplysia myoglobin does not contain cysteine.) Table 3 shows the partial sequences of the methionine-containing tryptic peptides TpIII and TpVII, as determined by means of CPA plus CPB digestion and dansyl-Edman degradation. TpVII contains two methionyl residues, the sequence of M_4CNBr and the amino terminal sequence of M_3CNBr, TpIII, on the other hand, starts with the two carboxyl terminal residues of M_3CNBr followed by the amino terminal sequence of M_2CNBr. The overall order of CNBr fragments therefore proves to be: M_1CNBr–M_4CNBr–M_3CNBr–M_2CNBr.

The sequence of M_3CNBr was determined by tryptic and chymotryptic digestion (table 4). The amino acid analysis (table 1) showed that the fragment contains 24 residues, among which one lysyl and one arginyl residue are present; it was therefore cleaved

Table 3
Amino acid sequence of methionine containing tryptic peptides of Aplysia myoglobin.

TpIII	→Ser–Met–Phe–Pro–Gly–Phe–Val–Ala (Asx, Ser, 2Pro, Gly, 5Ala, Val) ←Trp–Thr–Lys
TpVII	→Met–Ser–Ala–Met–Leu–Ser (Glx, Ala, Phe) Lys

→ determination by dansyl-Edman degradation; ← determination by carboxypeptidase digestion

Table 4
Carboxyl terminal sequence of Aplysia myoglobin.

Fragment	Peptide	Sequence	No. of residues
M_1CNBr		Acetyl x (8Lys, 3Arg, 15Asx, Thr, 3Glx, 3Pro, 4Gly, 18Ala, 6Val, 2Ile, 6Leu, 11Phe + Trp)–Lys(90)–Met(91)	91
M_4CNBr		Ser(92)–Ala(93)–Met(94)	3
M_3CNBr	Tp_2M_3	Leu(95)–Ser(96)–Gln(97)–Phe(98)–Ala(99)–Lys(100)	6
	Tp_3M_3	Glu(101)–His(102)–Val(103)–Gly(104)–Phe(105)–Gly(106)–Val(107)–Gly(108)–Ser(109)–Ala(110)–Gln(111)–Phe(112)–Glu(113)–Asn(114)–Val(115)–Arg(116) $Ch_3Tp_3M_3$ $Ch_1Tp_3M_3$ $Ch_2Tp_3M_3$	16
	Tp_1M_3	Ser(117)–Met(118)	2
M_2CNBr	Ch_3M_2	Phe(119)–Pro(120)–Gly(121)–Phe(122) LAP	4
	Ch_1M_2	Val(123)–Ala(124)–Ser(125)–Val(126)–Ala(127)–Ala(128)–Pro(129)–Pro(130)–Ala(131)–Gly(132)–Ala(133)–Asp(134)–Ala(135)–Trp(136)	14
	Ch_4M_2	Thr(137)–Lys(138)–Leu(139)–Phe(140)	4
	Ch_2M_2	Gly(141)–Leu(142)–Ile(143)–Ile(144)–Asp(145)–Ala(146)–Leu(147)	7
	Ch_5M_2	Lys(148)–Ala(149)–Ala(150)–Gly(151)–Lys(152)	5

→Determination by dansyl Edman degradation; ←Determination by carboxypeptidase digestion; LAP = Leucine Aminopeptidase digestion.

with trypsin into three peptides – Tp_1M_3, Tp_2M_3 and Tp_3M_3 – which were separated by chromatography on Dowex 50 × 2 columns in the experimental conditions described above. Dansyl-Edman degradation of the single peptides showed that the amino terminal sequence of Tp_2M_3 is Leu–Ser, identical to that of the M_3CNBr fragment. The presence of homoserine (methionine) in peptide Tp_1M_3 therefore imposed the order Tp_2M_3–Tp_3M_3–Tp_1M_3. It was however impossible to establish at this stage the complete sequence of Tp_3M_3; therefore this peptide was submitted to chymotryptic digestion. The resulting chymotryptic peptides, $Ch_1Tp_3M_3$, $Ch_2Tp_3M_3$ and $Ch_3Tp_3M_3$, were separated by Dowex 50 × 2 chromatography as described above. The known amino acid sequence of residues 1 to 7 of the Tp_3M_3 tryptic peptide proved the order of chymotryptic peptides to be $Ch_3Tp_3M_3$–$Ch_1Tp_3M_3$–$Ch_2Tp_3M_3$; all the more so because the carboxyl terminal residue of $Ch_2Tp_3M_3$ is arginine.

The sequence of M_2CNBr was established by means of chymotryptic and leucine aminopeptidase (LAP) digestion [14]. Chymotryptic hydrolysis cleaved the CNBr fragment into five chymotryptic peptides: Ch_1M_2, Ch_2M_2, Ch_3M_2, Ch_4M_2 and Ch_5M_2 (table 4), which were resolved by Dowex 50 × 2 chromatography in the experimental conditions described above. Dansyl-Edman degradation resulted in the determination of the complete sequences of all the chymotryptic peptides with the exception of Ch_1M_2; the latter is composed of 14 residues, only nine of which have been positioned by dansyl-Edman procedure. In order to complete the sequence, the peptide was subjected to LAP digestion, which yielded a nine-residue fragment (table 4). The LAP fragment was separated from the other reaction products by gel-filtration on a Sephadex G-10, 1 × 100 cm, column equilibrated with 0.2 M acetic acid, and its nine-residues sequence determined by the dansyl-Edman method. The known amino terminal sequence of M_2CNBr (table 2) made it possible to position Ch_3M_2 and Ch_1M_2 as the first and second chymotryptic peptides, respectively, at the amino terminal end of M_2CNBr. On the other hand the carboxyl terminal sequence of TpIII (table 3) indicates Ch_4M_2 as the peptide immediately following. Finally the carboxyl terminal sequence of Ch_5M_2 corresponds to that of the whole fragment M_2CNBr; the overall order of M_2CNBr chymotryptic peptides was thus established to be: Ch_3M_2–Ch_1M_2–Ch_4M_2–Ch_2M_2–Ch_5M_2 (table 4).

The glutamyl and asparagyl residues were identified with the aid of leucine aminopeptidase and pronase [15] digestion. The identification of the aspartyl residue in the Ch_1M_2 chymotryptic peptide was somewhat difficult. The nine-residue LAP fragment isolated as described above, was therefore submitted to dansyl-Edman degradation for the first three steps in order to remove proline. LAP digestion of the resulting hexapeptide proved the presence of an aspartyl residue in its fourth position.

Knowledge concerning the primary structure of Apylsia myoglobin is summarized in table 4, which shows both the relative position of the various peptides and the deduced sequences.

The unique chemical feature of Aplysia myoglobins, as previously remarked [2–4], is the presence of a single histidine per molecule, which presumably plays the role of the proximal histidine. In the partial sequence of Aplysia myoglobin the histidine corresponds to residue 102, the final assignment depending of course on the exact number of residues contained in the big M_1CNBr fragment. However its position is compatible with that normally occupied by the proximal histidine in hemoglobin and myoglobin, in which it occupies the positions: human alpha chain, 87; human beta chain, 92; sperm whale myoglobin, 93 [16]. It is interesting to note that in *Chironomus thummi thummi* hemoglobin the distal histidine is substituted by isoleucine [17, 18], showing that no unique role in O_2 binding can be attributed to the distal imidazole.

At this stage any detailed comparison of the primary structure of Aplysia myoglobin wuth that of other heme-proteins appears premature. Therefore considerations involving the evolutionary aspects of the problem, as well as possible implications of the structure on the functional properties, have to be postponed until the complete sequence of the protein, and possibly its tridimensional structure, are available.

Acknowledgements

The authors are greatly indebted to Mr. G.Franco for excellent technical assistance, and to Dr. M.Orlando for quantitative analysis of acetyl-groups.

References

[1] A.Rossi Fanelli and E.Antonini, Biokhimiya 22 (1957) 336.
[2] A.Rossi Fanelli, E.Antonini and D.Povoledo, in IUPAC Symp. on Protein Structure ed. A.Neubereger (Methuen, London, 1958).
[3] B.A.Wittenberg, M.Brunori, E.Antonini, J.B.Wittenberg and J.Wyman Arch. Biochem. Biophys. 111 (1965) 576.
[4] L.Tentori, G.Vivaldi, S.Carta, E.Antonini and M.Brunori, Nature 219 (1968) 487.
[5] D.N.Ward, J.A.Coffy, D.B.Ray and W.M.Lamkin, Anal. Chem. 14 (1966) 243.
[6] E.Gross and B.Witkop, J. Biol. Chem. 237 (1962) 1856.
[7] D.H.Spackman, W.H.Stein and S.Moore, Anal. Chem. 30 (1958) 1190.
[8] R.P.Ambler, in: Methods in Enzymology, Vol. XI, ed. S.P.Kolowick and N.O.Kaplan (Academic Press, New York, 1967) p. 161.
[9] W.R.Gray and B.S.Hartley, Biochem. J. 89 (1963) 379.
[10] W.R.Gray, in: Methods of Enzymology, Vol. XI, eds. S.P.Kolowick and N.O.Kaplan (Academic Press, New York, 1967) p. 469.
[11] K.R.Woods and K.T.Weng, Biochim. Biophys. Acta 133 (1967) 369.
[12] R.T.Jones, Cold Spring Harbor Symp. Quant. Biol. 29 (1964) 297.
[13] G.Toennies and J.Kolb, Anal. Chem. 23 (1951) 823.
[14] W.A.Schroeder, J.R.Shelton, J.B.Shelton, J.Cormik and R.T.Jones, Biochemistry 2 (1963) 992.
[15] J.A.Rothfus and E.L.Smith, J. Biol. Chem. 240 (1965) 4277.
[16] M.O.Dayhoff, Atlas of Protein Sequences (Natl. Biomed. Res. Found., Silver Spring, 1969).
[17] R.Huber, O.Hepp, H.Formanek and W.Steigmann, in press.