THE AMINO ACID SEQUENCE OF A DIMERIC MYOGLOBIN FROM THE GASTROPOD MOLLUSC, *BUSYCON CANALICULATUM* L

Alex G. BONNER and Richard A. LAURSEN
Department of Chemistry, Boston University, Boston, Massachusetts 02215, USA

Received 26 November 1976

1. Introduction

Several years ago Read et al. [1–6] described the isolation and characterization of the radular muscle myoglobins from several gastropod and amphineuran molluscs. All of these proteins are dimeric, with molecular weights of about 34 000; those from *Busycon canaliculatum* and *Buccinum undatum* are particularly interesting in that they show cooperative binding of oxygen (Hill coefficients of $n \geqslant 1.1$ and $n = 1.4$, respectively [1]). Since these proteins occupy a position between the tetrameric hemoglobins of vertebrates and the monomeric myoglobins, we felt it would be of value to investigate their primary structures as a first step towards understanding how they bind oxygen. The amino acid sequence of the *Busycon* myoglobin was determined and found to be homologous with, but about 80% different from, all other related globins of known sequence.

2. Materials and methods

Specimens of *Busycon canaliculatum* L. were purchased from the Marine Biology Laboratories, Woods Hole, Massachusetts and the myoglobin was isolated from the radular muscle according to Read [2]. Heme was removed [7] from the myoglobin and the apoprotein was either carboxymethylated to block the two cysteine residues or oxidized with performic acid.

The carboxymethylated protein was cleaved with cyanogen bromide in 70% HCOOH, and the peptides separated on Sephadex G-50 in 25% acetic acid. Tryptic and chymotryptic cleavage of oxidized or carboxymethylated protein was done in NH_4HCO_3 and peptides were purified by chromatography on the short column of the amino acid analyzer and by thin-layer chromatography. In some cases, the protein or peptides were maleylated and cleaved with trypsin at arginine.

Peptides were analyzed on a BioCal BC-200 amino acid analyzer and were sequenced by solid-phase Edman [8,9] degradation on a Sequemat Model 12 sequencer. Lysine peptides were coupled to sequencing resins by the di-isothiocyanate method [10]. Cyanogen bromide peptides were coupled at homoserine. Phenylthiohydantoins were identified by thin-layer chromatography and by HI hydrolysis to amino acids.

3. Results and discussion

Cyanogen bromide cleavage of *Busycon* myoglobin produced 6 peptides containing 1,6,20,24,32 and 64 residues, which, except for the largest peptide, could be sequenced directly by solid-phase Edman degradation. Overlaps were established using tryptic and chymotryptic peptides (fig.1). Peptide T16 was very difficult to isolate in pure form and thus the sequence G14–H3 (fig.1) must be regarded as tentative, although it is probably correct.

Busycon myoglobin is a typical globin in that it contains the invariant residues Pro C2, Phe CD1 and His F8 and a number of other residues, including the distal His E7, common to most globins [12]. In other myoglobins [13] a number of hydrophobic residues are found in the interior of the molecule lining the heme binding crevice. A similar pattern is noted for *Busycon* myoglobin and therefore it seems likely that

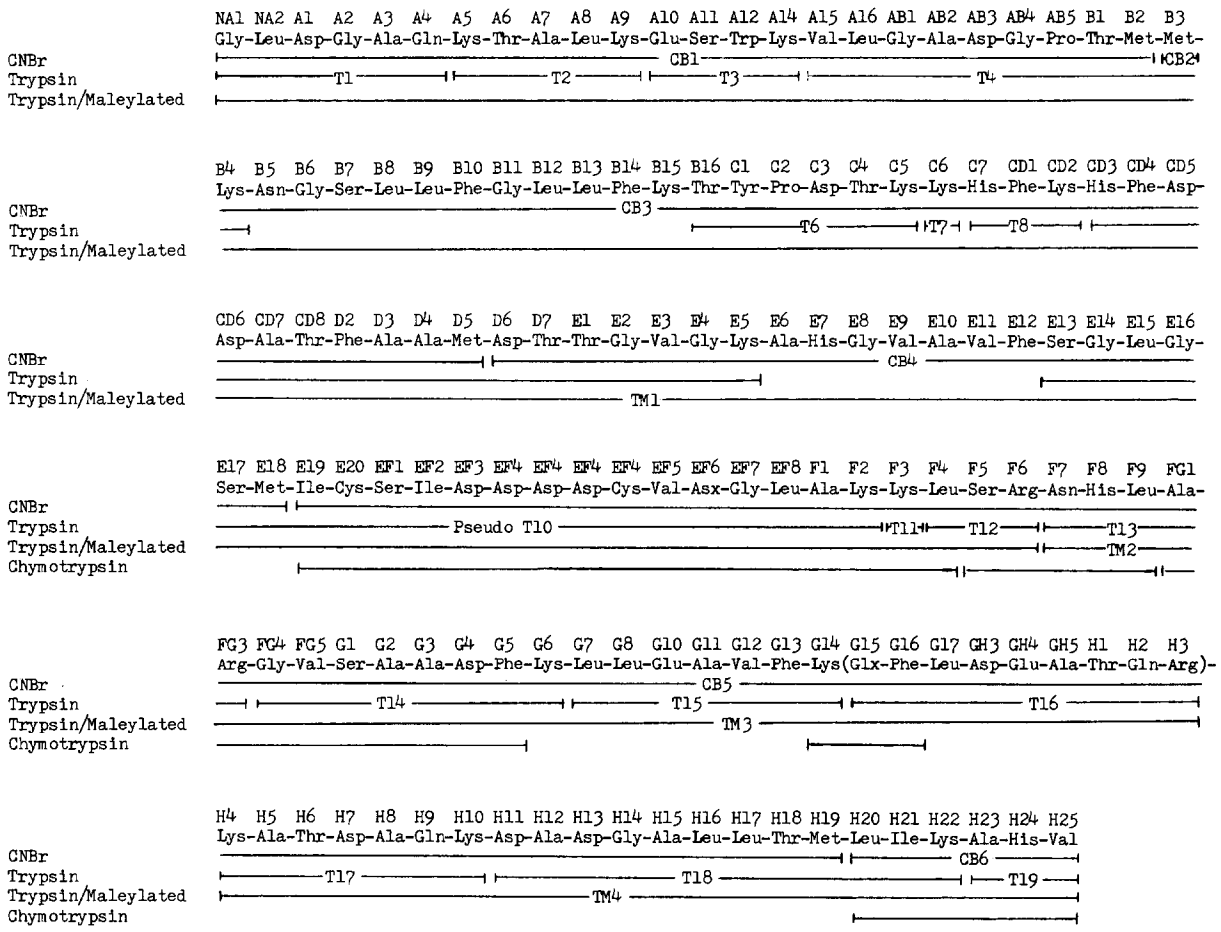

Fig.1. Amino acid sequence of *Busycon canaliculatum* myoglobin. The numbering of residues follows that for myoglobin (cf. ref. [13]). Deletions occur at A13, D1, FG2, G9 and G18–GH2; insertions are at AB2–5 and EF4. All tryptic and cyanogen bromide peptides, except CB5, were sequenced entirely. Portions of CB5, the maleylated tryptic peptides and some of the chymotryptic peptides were also sequenced. The hydrophobic peptide corresponding to T5 was not found (probably because of binding to Aminex resin) nor was the N-terminal portion of T10. T16 was difficult to purify in quantity and thus its sequence is tentative. A peptide corresponding to T17 minus the N-terminal lysine was also isolated.

the folding of this protein resembles that of other globins. Two points of difference are an insertion of four residues between the A and B helices and a deletion of five residues at the end of the G helix.

Although *Busycon* myoglobin is dimeric and shows cooperative binding of oxygen [1] we have not been able to discern any obvious similarity with the residues in the subunit contact regions (either $\alpha_1\beta_1$ or $\alpha_1\beta_2$) of hemoglobins [14]. Apparently the subunit interaction occurs in a different region or involves interaction of other amino acids.

The myoglobin from *Busycon canaliculatum* is 75–90% different from other globins* [12] that have been sequenced. It is interesting, though perhaps not significant, that the *Busycon* protein is more similar in sequence to human hemoglobin β-chain than to the myoglobin from the gastropod mollusc, *Aplysia limacina* (75% and 84% difference, respectively).

*We have completed half of the sequence from the more closely related *Buccinum undatum* myoglobin and find it to be only about 20% different from the *Busycon* protein.

Dayhoff and Barber [15] have suggested that the hemoglobins, which differ from the myoglobins by about 77%, diverged over a billion years ago. Assuming the same rate of mutation for molluscs, *Aplysia* and *Busycon* also must have diverged that long ago or earlier. However this estimate is at variance with the fossil record, which shows no precambrian (older than 600 million years) molluscan fossils, even though gastropods are among the oldest known molluscs [16]. Thus either the chronology of the fossil record or the estimated mutation rate for globins is in error.

The dimeric molluscan myoglobins, which show Hill coefficients ranging from none at all to 1.4 in *Buccinum undatum* [1] may provide a unique opportunity to study the evolution of cooperativity in a class of proteins. It is probable, because of their early date of divergence, that molluscs evolved cooperativity independently of the vertebrates. It would be of interest to learn whether molluscs arrived at the same solution as the vertebrates. The answer to this question will require a crystallographic study.

Acknowledgements

This work was supported by grants from the National Science Foundation. R.A.L. was the recipient of a NIH Research Career Development Award and an Alfred P. Sloan Fellowship.

References

[1] Terwilliger, R. C. and Read, K. R. H. (1971) Int. J. Biochem. 2, 253–261.
[2] Read, K. R. H. (1966) Comp. Biochem. Physiol. 17, 375–390.
[3] Terwilliger, R. C. and Read, K. R. H. (1970) Int. J. Biochem. 1, 281–291.
[4] Read, K. R. H. (1968) Comp. Biochem. Physiol. 25, 81–94.
[5] Johnson, J. P., Koppenheffer, T. L. and Read, K. R. H. (1971) Int. J. Biochem. 2, 65–71.
[6] Johnson, J. P. and Read, K. R. H. (1971) Int. J. Biochem. 2, 293–306.
[7] Teale, F. W. J. (1959) Biochim. Biophys. Acta 35, 543–548.
[8] Laursen, R. A. (1971) Eur. J. Biochem. 20, 89–102.
[9] Laursen, R. A. (ed) (1975) Solid-Phase Methods in Protein Sequence Analysis, Pierce Chemical Co., Rockford, Ill.
[10] Laursen, R. A., Horn, M. J. and Bonner, A. G. (1972) FEBS Lett. 21, 67–70.
[11] Horn, M. J. and Laursen, R. A. (1973) FEBS Lett. 36, 285–288.
[12] Hunt, L. T. and Dayhoff, M. O. (1976) in: Atlas of Protein Sequence (Dayhoff, M. O. ed) Vol. 5, Suppl. 2, pp. 191–224, National Biomedical Research Foundation, Silver Spring, Md.
[13] Romero-Herrera, A. E., Lehmann, H., Joysey, K. A. and Friday, A. E. (1973) Nature 246, 389–395.
[14] Perutz, M. F. (1970) Nature 228, 726–734.
[15] Dayhoff, M. O. and Barker, W. C. (1972) in: Atlas of Protein Sequence (Dayhoff, M. O. ed) Vol. 5, pp. 41–45, National Biomedical Research Foundation, Silver Spring, Md.
[16] Solem, A. (1974) The Shell Makers. Introducing Mollusks, p. 40. Wiley, New York.