THE AMINO ACID SEQUENCE OF CYTOCHROME *c* FROM *ASTERIAS RUBENS* L. (COMMON STARFISH)

Andrew LYDDIATT* and Donald BOULTER
Department of Botany, University of Durham, Durham DH1 3LE, England

Received 10 April 1976
Original figures received 5 June 1976

1. Introduction

The large majority of amino acid sequence data for cytochrome *c* are from higher plant and animal sources, and of the few known invertebrate sequences, the majority derive from the insects [1,2]. As part of a study of the molecular evolution of invertebrate cytochrome *c*, it was necessary to determine the feasibility of obtaining representative data throughout this group of animals. This paper reports the purification and determination of the amino acid sequence of cytochrome *c* from the echinoderm *Asterias rubens* L.

2. Materials and methods

Starfish were collected from the estuary at Burnham-on-Crouch and frozen immediately. They were stored at −20°C until required. All other materials were as previously described [3–5].

Cytochrome *c* was isolated and purified as previously described [5–7] with the exception that the homogenisation and extraction steps were performed at pH 6.5.

The amino acid sequence was determined by the dansyl-phenylisothiocyanate method as previously described [3–5].

* Present address: Department of Genetics, Trinity College, Dublin 2, Eire.

3. Results

A total of 24 mg cytochrome *c*, with an absorption ratio $E^{C3+}_{410}/E^{C3+}_{280}$ equal to 4.0, was purified from 100 kg of *Asterias*. The protein did not precipitate in 60–100% saturated solutions of ammonium sulphate. N-terminal analysis of the total protein by the method of [8,9] gave a single residue glycine, as the N-terminus. The amino acid composition was determined from three duplicate 50 nmol samples of protein hydrolysed for 24, 48 and 72 h, respectively. The values obtained for serine, glutamic acid and lysine did not agree with the sequence results, for which there was redundant evidence.

The amino acid sequence was deduced from the sequence analysis of overlapping peptides isolated from chymotryptic and tryptic peptides, and this is given in fig.1. All residues were positively identified by sequence analysis in both chymotryptic and tryptic digestions, with the exception of residues 16, 17, 22, 37, 38, 53, 70–72, which were placed from the evidence of one digestion and peptide amino acid composition data. All overlaps between chymotryptic and tryptic peptides were observed with the exception of the region C2CA-C2CB/T4CA-T4CB where the order was apparent from a consideration of other cytochromes *c* [1,2], and the region C3B-C4/T7-T8. Here the placement of peptides was unambiguous from an inspection of the sequence data for C3B, C4, C4CA, T7 and T8. All amides indicated in the sequence were placed from

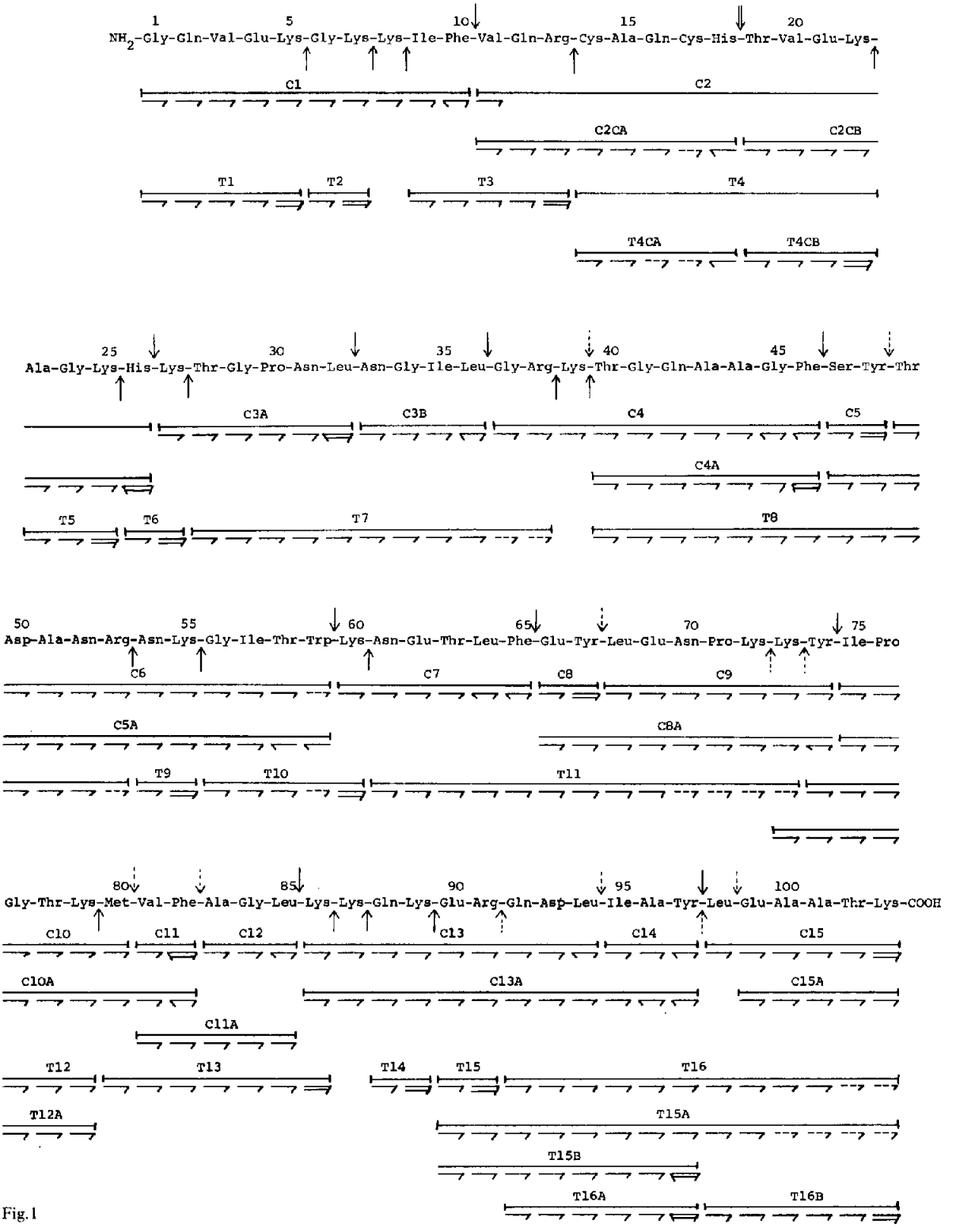
1 5 10 15 20
NH2-Gly-Gln-Val-Glu-Lys-Gly-Lys-Lys-Ile-Phe-Val-Gln-Arg-Cys-Ala-Gln-Cys-His-Thr-Val-Glu-Lys-
C1
C2
C2CA
C2CB
T1
T2
T3
T4
T4CA
T4CB
25 30 35 40 45
Ala-Gly-Lys-His-Lys-Thr-Gly-Pro-Asn-Leu-Asn-Gly-Ile-Leu-Gly-Arg-Lys-Thr-Gly-Gln-Ala-Ala-Gly-Phe-Ser-Tyr-Thr
C3A
C3B
C4
C5
C4A
T5
T6
T7
T8
50 55 60 65 70 75
Asp-Ala-Asn-Arg-Asn-Lys-Gly-Ile-Thr-Trp-Lys-Asn-Glu-Thr-Leu-Phe-Glu-Tyr-Leu-Glu-Asn-Pro-Lys-Lys-Tyr-Ile-Pro
C6
C7
C8
C9
C5A
C8A
T9
T10
T11
80 85 90 95 100
Gly-Thr-Lys-Met-Val-Phe-Ala-Gly-Leu-Lys-Lys-Gln-Lys-Glu-Arg-Gln-Asp-Leu-Ile-Ala-Tyr-Leu-Glu-Ala-Ala-Thr-Lys-COOH
C10
C11
C12
C13
C14
C15
C10A
C13A
C15A
C11A
T12
T13
T14
T15
T16
T12A
T15A
T15B
T16A
T16B

Fig. 1

the electrophoretic mobilities at pH 6.5 of intact or partially degraded peptides [10]. The observed enzyme specificities were consistent with those expected [11,12] except that partial chymotryptic cleavage was observed at lysine-39 and partial tryptic cleavage at tyrosine-97.

4. Discussion

The isolation of peptide C4A in low yield, indicating partial chymotryptic activity at lysine-39 is unusual although to predict the course of chymotryptic hydrolysis at non-aromatic residues is difficult [12]. Many other cytochromes *c* have an identical sequence (–Gly–Arg–Lys–Thr–Gly–) to *Asterias* in the region 37–41, but no similar anomalous cleavage has been reported [1,2], although partial hydrolysis at threonine-40 was reported in human [13]. However, the activity of TPCK-trypsin at tyrosine-97 of cytochrome *c* is quite common and may be due to a less than complete ethanolic denaturation coupled with the position of this residue in the three-dimensional structure of the molecule [14,15].

Asterias cytochrome *c* consists of a single polypeptide chain 103 residues in length and is homologous with other cytochromes *c* when arranged in the standard alignment [1]. It is the first echinoderm sequence to be determined and like the *Helix* sequence [16] possesses a non-acetylated, N-terminal glycine in position-1 of the alignment. Of the remaining invertebrates, the insect sequences [1,5,17] and the annelid sequence [18] all have a non-acetylated tail of 4 or 5 residues at the N-terminus, whilst the crustacean has no such tail and carries a blocked N-terminus [19].

The amino acid sequence of *Asterias* cytochrome *c*, together with other invertebrate sequences and representative sequences from other taxonomic groups [1,2,5,16–20] was used to construct a molecular phylogeny relating 21 sequences using an approach based upon the ancestral sequence method [1] (fig.2). Two equal alternatives were found for the position of bullfrog cytochrome *c* with an identical

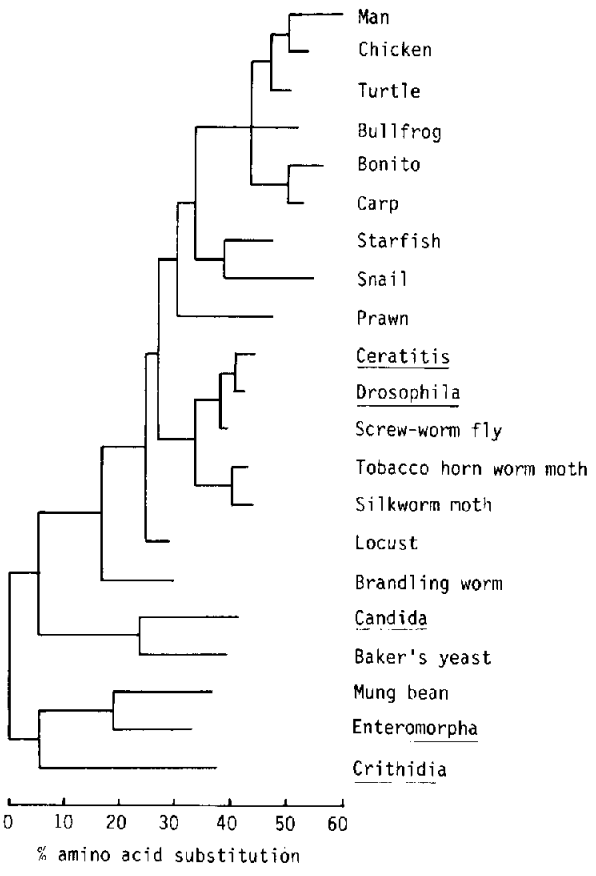

Fig.2. A phylogeny constructed by the ancestral sequence method relating 21 amino acid sequences of cytochrome *c*. Sequences were aligned relative to the cysteinyl residue; absent residues at the N-terminus and C-terminus were considered as deletions and computed as amino acid differences. The position of bullfrog is summarised (see text).

Fig.1. The amino acid sequence of *Asterias* cytochrome *c*. Residues were identified by dansyl-phenylisothiocyanate analysis (——), amino acid composition data (- - -), carboxypeptidase-A analysis (——) and as the free C-terminal amino acid following the final Edman degradation step (⟹). Arrows (↑↓, ↑↓ and ⇓) indicate points of complete, partial and secondary cleavage respectively, up for trypsin and down for chymotrypsin. Prefixed T and C refer to tryptic and chymotryptic peptides respectively, and peptides derived from partial cleavage have a letter subscript to the major peptide. Peptides C2 and T4 were digested with chymotrypsin following the removal of the heme moiety.

minimum number of amino acid substitutions required to relate the overall tree. No equal alternative was found for the invertebrate region of the phylogeny.

The position of *Asterias* sharing a more recent ancestral sequence with the vertebrates than that shared between the insects and vertebrates agrees with the classical view of phylogeny [21,22]. However, the starfish-snail grouping is remarkable, for the molluscs are generally thought to have diverged from a protostomial line of animal descent and thus to be more closely related to the annelids and arthropods [23,24]. The molecular phylogeny provides no support for the embryologically based Protostomia–Deuterostomia division of the animal kingdom [21–24].

Errors in molecular phylogenies have been calculated to increase with increasing branch-lengths between nodes and present-day sequences [25]. Any errors associated with the mollusc–echinoderm region of fig.2 will be greater than those associated with the insect grouping as a result of the number of sequences known for the latter. More data is required to shorten branch-lengths for the under-represented phyla, and to subsequently increase the resolution of the overall phylogeny.

Acknowledgments

We are grateful to Dr D. Peacock for the computer analysis and to Dr P. S. Jefferies for helpful discussions. This research was supported by a grant from the Science Research Council.

References

[1] Dayhoff, M. O. (1972) Atlas of Protein Sequence and Structure, Vol. 5, National Biomedical Research Foundation, Silver Spring, Maryland, USA.
[2] Dayhoff, M. O. (1973) Atlas of Protein Sequence and Structure, Vol. 5 Supplement, National Biomedical Research Foundation, Silver Spring, Maryland, USA.
[3] Thompson, E. W., Laycock, M. V., Ramshaw, J. A. M. and Boulter, D. (1970) Biochem. J. 117, 183–192.
[4] Thompson, E. W., Richardson, M. and Boulter, D. (1971) Biochem. J. 121, 439–446.
[5] Lyddiatt, A. Ph. D. Thesis, University of Durham.
[6] Richardson, M., Laycock, M. V., Ramshaw, J. A. M., Thompson, E. W. and Boulter, D. (1970) Phytochem. 8. 2271–2280.
[7] Richardson, M., Richardson, D., Ramshaw, J. A. M., Thompson, E. W. and Boulter, D. (1971) J. Biochem. 69, 811–813.
[8] Gros, C. and Labouesse, B. (1969) Eur. J. Biochem. 7, 463–470.
[9] Gray, W. R. (1967) in: Methods in Enzymology (Colowick, S. P. and Kaplan, N. O. eds.) Vol. 11, pp. 139–151 Academic Press, New York and London.
[10] Offord, R. E. (1966) Nature, 211, 591–593.
[11] Smyth, D. G. (1967) in: Methods in Enzymology (Colowick, S. P. and Kaplan, N. O. eds.) Vol. 11, pp. 214–231 Academic Press, New York and London.
[12] Kasper, C. B. (1970) in: Protein Sequence Determination, Molec. Biol. Biochem. and Biophys., Vol. 8 (Needleman, S. B. eds.) pp. 137–184 Chapman and Hall, London.
[13] Matsubara, H. and Smith, E. L. (1963) J. Biol. Chem. 238, 2732–2753.
[14] Dickerson, R. E. (1971) J. Molec. Evol. 1, 26–45.
[15] Boulter, D. and Ramshaw, J. A. M. (1972) Phytochem. 11, 553–561.
[16] Brown, R. H. Richardson, M., Boulter, D., Ramshaw, J. A. M. and Jefferies, R. P. S. (1972) Biochem. J. 128, 971–974.
[17] Fernandez-Sousa, J. M., Gavilanes, J. G., Municio, A. M., Paredes, J. A., Perez-Aranda, A. and Rodriguez, R. (1975) Biochem. Biophys. Acta 393, 358–367.
[18] Lyddiatt, A. and Boulter, D. (1976) FEBS Lett. 62, 85–88.
[19] Lyddiatt, A. and Boulter, D. (1976) Comp. Biochem. Physiol. in press.
[20] Meatyard, B. T. and Boulter, D. (1974) Phytochem. 13, 2777–2782.
[21] Grobben, K. (1908) Verh. Zool. Bot Ges Wien 58, 491–505.
[22] Marcus, E. (1958) Quart. Rev. Biol. 33, 24–58.
[23] Hyman, L. H. (1940–1959) The Invertebrates, Vol. 1–5, McGraw Hill, N.Y.
[24] Hanson, E. D. (1961) Animal Diversity, Prentice-Hall, New Jersey.
[25] Peacock, D. and Boulter, D. (1957) J. Molec. Biol. 95, 513–527.