THE AMINO ACID SEQUENCE OF CYTOCHROME *c* FROM *EISENIA FOETIDA* (SAVIGNY) (COMMON BRANDLING WORM)

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

Received 20 October 1975

1. Introduction

Amino acid sequences of cytochrome *c* have been applied extensively in the development and evaluation of strategies for the construction of molecular phylogenies [1–3]. The data accumulated shows a strong bias towards vertebrate and higher plant sources [4,5]. The present paper reports on the primary structure of cytochrome *c* of *Eisenia foetida* (common brandling worm), as part of a study of the molecular evolution of this protein from invertebrates.

2. Materials and methods

Brandling worms were obtained alive from Lime Tree Worm Farm, Bullwell, Nottingham, and used immediately. All other materials were as previously described [6–8].

Cytochrome *c* was isolated and purified as previously described [9,10] with the following modifications. (i) Fresh worms were homogenised (25 kg batches) with 1 vol of ice in 2 vol 10 mM $Al_2(SO_4)_3$ maintained at pH 4.5 with 2 M H_2SO_4, and stirred for 2 h at 2°C. Following centrifugations at pH 4.5 and pH 8.0, the supernatant (90 litres) was pumped through an 18 X 10 cm column of Amberlite CG-50 at 15 litres per h and cytochrome *c* was eluted batchwise in 2 M NaCl maintained at pH 8.0 with 2 M NaOH. (ii) *Eisenia* cytochrome *c* showed no precipitation in 60–100% saturated solutions of ammonium sulphate. 80% saturated solutions were centrifuged at 30 000 *g* (r_{av} 4.2 cm) for 20 min at 2°C, the supernatant was exhaustively dialysed against 10 mM NaH_2PO_4–Na_2HPO_4, pH 7.2, and the cytochrome *c* concentrated on a 1 X 5 cm column of CM-52 cellulose. (iii) Following gel filtration on a 2 X 90 cm column of Biogel P-30, the purest fractions were adsorbed onto a 1 X 30 cm column of CM-52 cellulose and eluted in a linear pH gradient from 10 mM NaH_2PO_4–NaH_2PO_4, pH 7.2, to 10 mM $Na_4P_2O_7$–NaOH, pH 11.5. Fractions were collected into 1.0 ml of 10 mM phosphate buffer, pH 7.2. The purest fractions were taken to pH 7.2 with 1 M HCl and the chromatography repeated using a linear ionic gradient from 10 mM to 300 mM NaH_2PO_4–Na_2HPO_4, pH 7.2.

The amino acid sequence was determined by the dansyl-phenylisothiocyanate method as previously described [6–8].

3. Results

A total of 26 mg cytochrome *c*, having an absorption ratio $E^{C2+}_{550}/E^{C3+}_{280}$, equal to 1.1, was purified from 75 kg of *Eisenia*. N-terminal analysis of the total protein [11,12] showed 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 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 digestions and this is shown in fig.1. All residues

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

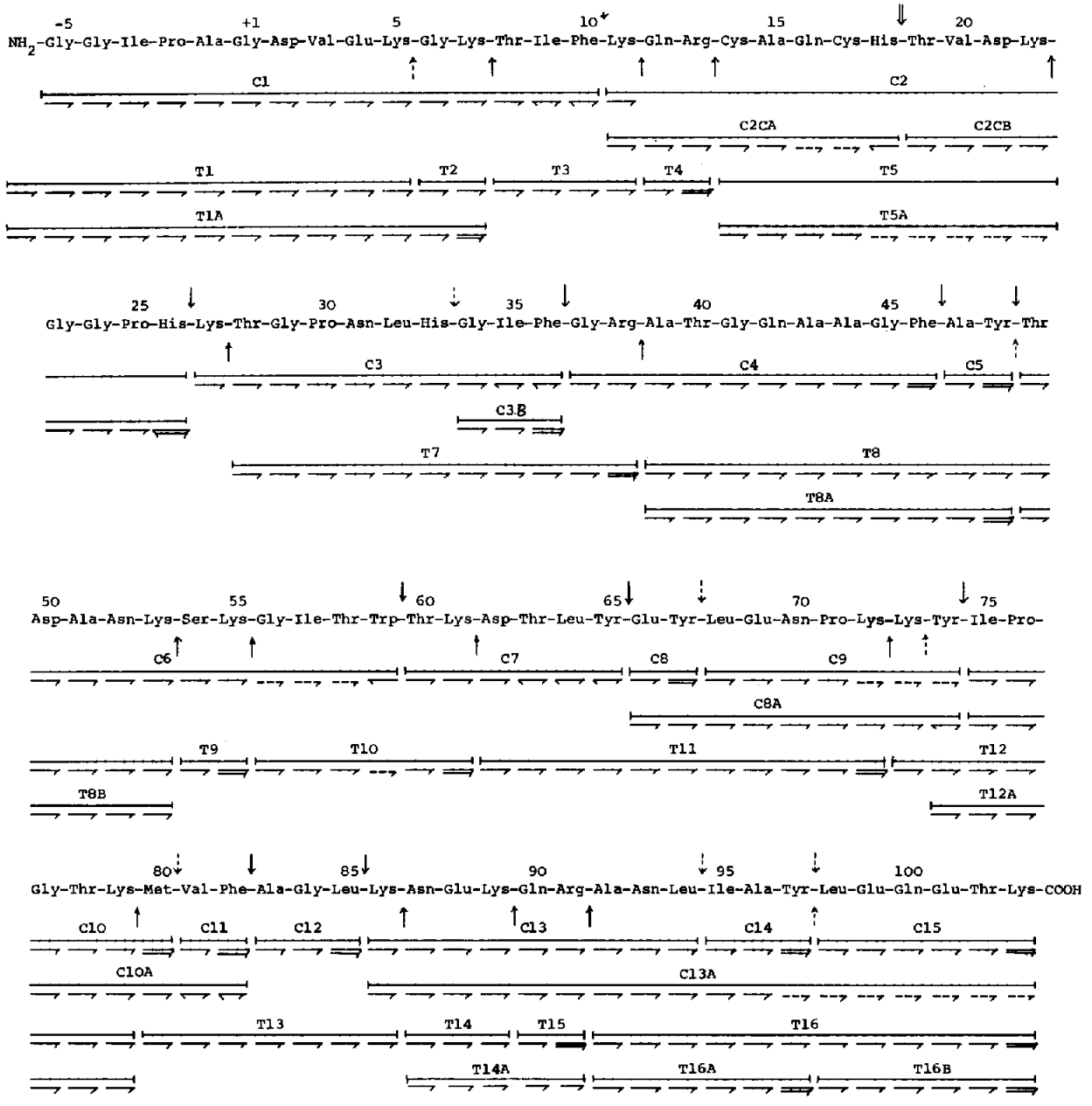

Fig.1. The amino acid sequence of *Eisenia* cytochrome *c*. Residues were identified by dansyl-phenylisothiocyanate analysis (——>), amino acid composition data (– – >), carboxypeptidase-A digestion (<——) and as the free C-terminal amino acid following the final Edman degradation (==>). Arrows (↑ ↓, ⇑ ⇓ and ⇓) indicate points of complete, partial and secondary enzyme cleavage; up for trypsin and down for chymotrypsin. Prefixes T and C refer to tryptic and chymotryptic peptides respectively, and peptides derived from partial cleavage have a letter subscript to the major peptide. Peptide C2 was digested with chymotrypsin after removal of the heme moiety.

were positively identified in both digestions with the exception of residues 16–27 and 56–59 which were placed from the evidence of one digestion and amino acid composition data. All the required overlaps between chymotryptic peptides were observed with the exception of the region C2CB-C3/T5A-T7 where, because of the non-isolation of T6, the order of peptides was deduced from a consideration of the appropriate region in other cytochromes *c* [4]. All acidic and amide residues were placed from the electrophoretic mobilities of intact or partially degraded peptides [13]. Observed enzyme specificities were consistent with those expected [14,15] except that partial tryptic cleavage was observed at tyrosine-48 and tyrosine-97.

4. Discussion

Eisenia cytochrome *c* consists of a single polypeptide, 108 residues in length, and is homologous with other cytochromes *c* when arranged in the standard alignment [4]. The molecule has at the N-terminal region a non-acetylated, five residue tail relative to glycine-1 of the standard alignment [4]. The tail (Gly–Gly–Ile–Pro–Ala) is very similar to the four residue tail of insect cytochrome *c* (Gly–Val–Pro–Ala or Gln) but less similar to the longer structures found in higher plants and some fungi. By comparison *Helix* (garden snail) cytochrome *c*, has no N-terminal tail [16].

The amino acid sequence of *Eisenia* cytochrome *c*, together with other invertebrate sequences [4,8,16,17] and representative sequences from other taxonomic groups [4,5,18] were used to construct a molecular phylogeny relating 18 sequences using computer programming based upon the ancestral sequence method [4,19]. This is shown in fig.2. Two equal alternatives were found for the position of bullfrog cytochrome *c*, both of which had the same minimum number of amino acid substitutions required to relate all the sequences of the tree. No equal alternatives were found for the invertebrate region of the topology. A computation of the 103 residue positions common to all the sequences produced an identical topology, demonstrating the insect and worm sequences to be related on similarities other than the N-terminal tail regions.

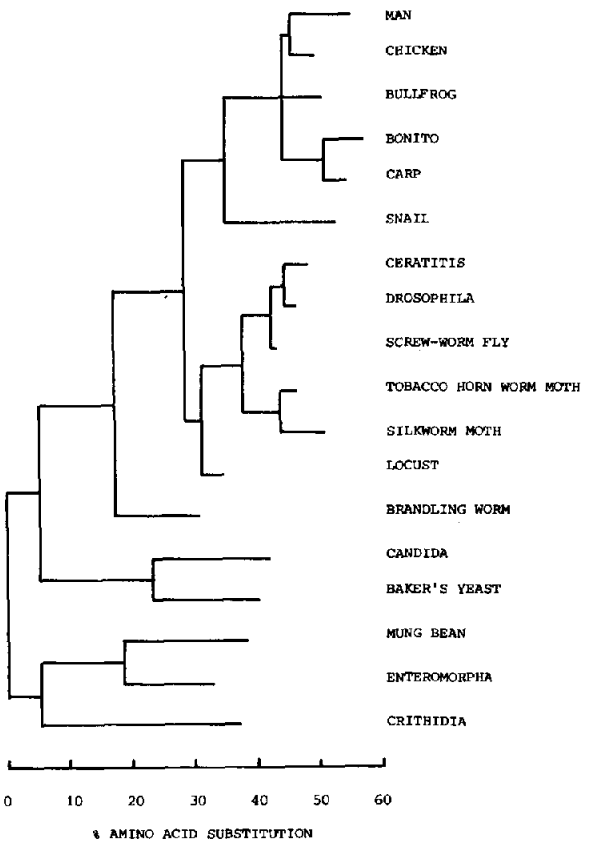

Fig.2. A phylogeny constructed by the ancestral sequence method relating 18 amino acid sequences of cytochrome *c*. Sequences were aligned relative to the cysteinyl residues; absent residues at the N-terminus and C-terminus were computed as amino acid differences. Two equal, alternative positions for bullfrog on the human-chicken and carp–bonito branches are summarized.

The indication that the insects shared a more recent common ancestral cytochrome *c* with the vertebrates than that shared between *Eisenia* (Annelida) and the vertebrates, agrees with the majority view of classical phylogeny [20–22]. However, the position of *Helix* (Mollusca) in fig.2 is less in agreement with classical ideas. However, further sequence data is required of mollusc and annelid cytochromes *c* in order to establish their position with certainty in a molecular phylogeny.

Acknowledgements

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

References

[1] Boulter, D., Ramshaw, J. A. M., Thompson, E. W., Richardson, M. and Brown, R. H. (1972) Proc. Roy. Soc. London B. 181, 441–445.
[2] McLaughlin, P. J. and Dayhoff, M. O. (1973) J. Molec. Evol. 2, 99–116.
[3] Langley, C. H. and Fitch, W. M. (1974) J. Molec. Evol. 3, 161–178.
[4] Dayhoff, M. O. (1972) Atlas of Protein Sequence and Structure, Vol. 5, National Biomedical Research Foundation, Silver Spring Maryland, USA.
[5] Dayhoff, M. O. (1973) Atlas of Protein Sequence and Structure, Vol. 5, Supplement, National Biomedical Research Foundation, Silver Spring, Maryland, USA.
[6] Thompson, E. W., Laycock, M. V., Ramshaw, J. A. M. and Boulter, D. (1970) Biochem. J. 117, 183–192.
[7] Thompson, E. W., Richardson, M. and Boulter, D. (1971) Biochem. J. 121, 439–446.
[8] Lyddiatt, A. and Boulter, D. (1975) Biochem. J. submitted for publication.
[9] Richardson, M., Laycock, M. V., Ramshaw, J. A. M., Thompson, E. W. and Boulter, D. (1970) Phytochem. 9, 2271–2280.
[10] Richardson, M., Richardson, D., Ramshaw, J. A. M., Thompson, E. W. and Boulter, D. (1971) J. Biochem. 69, 811–813.
[11] Gray, W. R. (1972) in: Methods in Enzymology (Colowick, S. P. and Kaplan, N. O., eds.) Vol. 11, 121–138, Academic Press, N. Y. and London.
[12] Gros, C. and Labouesse, B. (1969) Eur. J. Biochem. 7, 463–470.
[13] Offord, R. E. (1966) Nature 211, 591–593.
[14] Smyth, D. G. (1967) in: Methods in Enzymology (Colowick, S. P. and Kaplan, N. O., eds.) Vol. 11, 214–231, Academic Press, N. Y. and London.
[15] Kasper, C. B. (1970) in: Protein Sequence Determination, Molec. Biol. Biochem. and Biophys., Vol. 8 (Needleman, S. B., ed.), 137–184. Chapman and Hall, London.
[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) Biochim. Biophys. Acta 393, 358–367.
[18] Meatyard, B. T. and Boulter, D. (1974) Phytochem. 13, 2777–2782.
[19] Dayhoff, M. O. and Eck, R. V. (1966) Atlas of Protein Sequence and Structure, Vol. 2, National Biomedical Research Foundation, Silver Spring, Maryland USA.
[20] Hyman, L. H. (1940–1959) The Invertebrates, Vols. 1–5. McGraw-Hill, N. Y.
[21] Hanson, E. D. (1961) Animal Diversity, Prentice-Hall, New Jersey.
[22] Marcus, E. (1958) Quart. Rev. Biol. 33, 24–58.