Biochem. J. (1972) **128**, 971–974
Printed in Great Britain

The Amino Acid Sequence of Cytochrome *c* from *Helix aspersa* Müller (Garden Snail)

By R. H. BROWN, M. RICHARDSON, D. BOULTER, J. A. M. RAMSHAW and R. P. S. JEFFERIES*
Department of Botany, University of Durham, Durham City, U.K.

(*Received* 10 *February* 1972)

The amino acid sequence of a snail cytochrome *c* has been determined. The molecule consists of a single polypeptide chain of 104 residues, and is homologous with other mitochondrial cytochromes *c*. Unlike the cytochromes *c* from vertebrates, there is no acetyl blocking group at the *N*-terminus. A change in an otherwise invariant position has been observed in position 87. Comparison with amino acid sequences of cytochromes *c* from other sources indicates that the point of divergence of the molluscs and the vertebrates in evolutionary time was 720 million years ago. Experimental details are given in a supplementary paper that has been deposited as Supplementary Publication SUP 50009 at the National Lending Library for Science and Technology, Boston Spa, Yorks. LS23 7BQ, U.K., from whom copies can be obtained on the terms indicated in *Biochem. J.* (1972), **126**, 5.

The amino acid sequences of a large number of cytochromes *c* have now been determined (see information collected by Dayhoff, 1969). The sources of these cytochromes include representatives of the vertebrates, insects, plants and fungi. It was of interest, therefore, to determine the sequence of a representative of the molluscs.

Materials

Live snails were bought from Leytons (Foods) Ltd., Priddy, Wells, Somerset, U.K., and were deep frozen until required. Their place of origin was the Mendip Hills. Other materials were as described by Thompson *et al.* (1971).

Methods

The isolation and purification of the cytochrome *c* from a snail (*Helix aspersa* Müller) were done essentially as described for plant material by Richardson *et al.* (1971). The cytochrome adsorbed to the Amberlite resin was eluted with 1M-NaCl instead of 0.5M-NaCl. The cytochrome *c* eluted from the Biogel P-30 column was considered sufficiently pure for the analysis of the sequence and was not subjected to further purification procedures. The preparation used for analysis was homogeneous on polyacrylamide gel electrophoresis, and had an E_{550} (reduced)/E_{280} (oxidized) ratio of 1.14. A 2μmol portion of the preparation was used in the analysis.

The methods used for the determination of the sequence were described by Thompson *et al.* (1971).

The *C*-terminal fragment of 24 residues after methionine was separated by gel filtration from the larger fragment, resulting from cyanogen bromide cleavage of the cytochrome. After digestion with chymotrypsin, the resulting peptides were analysed by the same methods used for the chymotryptic digest of whole cytochrome.

Results

The amino acid compositions of *H. aspersa* cytochrome *c*, determined by analysis and calculated from the sequence, are shown in Table 1. These values are in reasonable agreement, although the values obtained by analysis for proline and arginine were higher than the number calculated from the sequence. Fig. 1 gives the complete sequence of *H. aspersa* cytochrome *c*, together with an indication of the chymotryptic and tryptic peptides identified.

Consideration of the mobilities of the non-haem peptides indicated that there were 11 amide groups. Of these, six can be placed unambiguously and two by homology, these latter being asparagine residues in positions 52 and 70. Position 90 can be placed as glutamine by homology with other known sequences, leaving unplaced single amide positions at residues 2 or 4, 87 or 88 and 69 or 70.

Comparison of the mobility of the chymotryptic haem peptides of snail and horse heart (Margoliash *et al.*, 1961) reveals that the snail peptide has a lower charge. Hence, the snail peptide contains one fewer basic residue, and positions 16 and 21 may be placed

* Department of Palaeontology, British Museum (Natural History), Cromwell Road, London S.W.7, U.K.

Table 1. *Amino acid composition of snail cytochrome c*

The 24 and 72h values are the average of two determinations. N.D., not determined.

Hydrolysis time ...	Amino acid composition (residues/molecule) 24h	72h	Average	Sequence
Asp	6.59	6.51	6.55	6
Thr	7.56	6.04	8.45*	9
Ser	1.88	1.49	2.07*	1
Glu	13.4	13.7	13.6	14
Pro	7.1	6.6	6.8	4
Gly	13.3	13.3	13.3	13
Ala	6.56	6.64	6.60	8
Val	3.33	3.37	3.35	3
Cys	2.06	1.69	1.88	2
Met	0.83	0.97	0.97†	1
Ile	4.57	4.37	4.47	4
Leu	6.98	7.13	7.06	8
Tyr	3.42	3.48	3.45	4
Phe	4.37	4.71	4.54	5
Lys	17.5	17.3	17.4	16
His	2.44	2.57	2.51	3
Arg	3.09	3.30	3.20	2
Trp	N.D.	N.D.		1
Total				104

* Extrapolated to zero time assuming first-order kinetics of destruction.

† Maximal value.

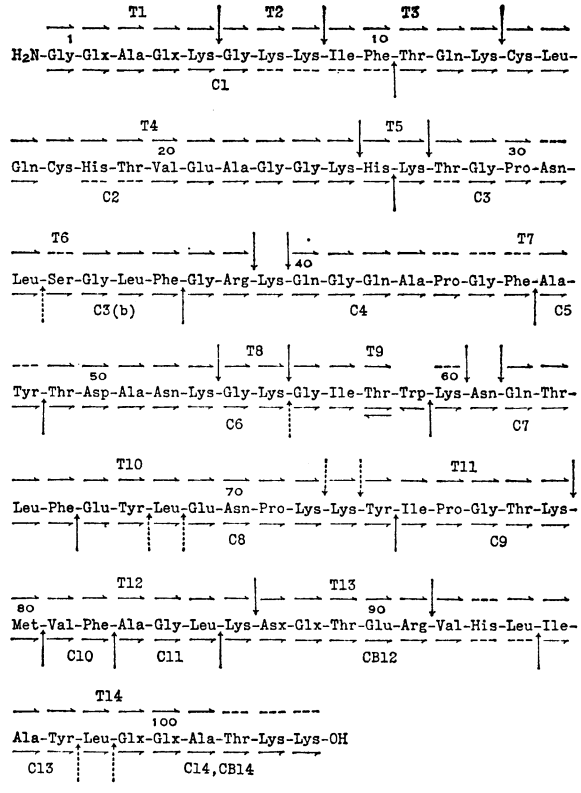

Fig. 1. *Amino acid sequence of snail cytochrome c*

---, Composition determined qualitatively; ——→, sequence determined by the dansyl–Edman method; ←——, sequence determined by using carboxypeptidase; ↑↓, major enzymic cleavages; ⇡⇣, partial enzymic cleavages. Abbreviations: T, tryptic peptide; C, chymotryptic peptide; CB, chymotryptic peptide after cyanogen bromide cleavage. For full details of individual peptides, see the Supplementary Publication (SUP 50009).

by homology as glutamine and glutamic acid respectively. The total number of amides is therefore 12, three of which cannot be placed on present evidence.

Unlike vertebrate cytochrome *c* sequences, snail cytochrome *c* has no acetyl blocking group at the *N*-terminus. Its absence was confirmed by using the dansyl–Edman method on the whole cytochrome, which successfully revealed the first five residues.

The *C*-terminal residue was confirmed as lysine by carboxypeptidase-A digestion of the whole cytochrome. The molecule thus has a total of 104 residues, in common with cytochrome *c* from vertebrates.

The partial cleavage of the Lys-Gly bond at positions 55–56 was the only evidence of tryptic-type activity in the chymotryptic digest. It was not possible to identify position 39 positively as lysine in the tryptic digest, since the sequence Arg-Lys-Gln in positions 38–40 gave rise to free lysine on tryptic digestion. Free lysine would also be expected to arise from the several Lys-Lys sequences in the molecule. Chymotryptic-type activity was observed in the tryptic digest in the apparent split between Asn-Gln at positions 61–62. Since position 60 is lysine, asparagine 61 would be released as a free amino acid on tryptic digestion.

Discussion

Snail cytochrome *c* contains 21 basic residues and 20 acidic residues, of which 12 are amides. It therefore has an excess of 13 basic over free acid residues. Horse-heart cytochrome *c* (Margoliash *et al.*, 1961) has a 'basic excess' of 12 residues. These values are consistent with the observed electrophoretic behaviour of the cytochromes (J. Valentine, unpublished results), snail cytochrome having a slightly higher mobility at pH8.7 than horse-heart cytochrome. In contrast, plant cytochromes *c* have lower 'basic excesses' of 6–8 residues and correspondingly lower mobilities. It is interesting that a higher concentration of salt was required to elute the snail cytochrome

from Amberlite than has been used for eluting plant cytochrome *c*.

A change in an otherwise invariant residue has been observed in position 87. Asparagine (or aspartic acid) replaces lysine in this position. Glutamic acid or glutamine in position two replaces an invariant aspartic acid in vertebrate sequences at present available, although variations have been observed in insects, yeasts and plants. Leucine in position 15 replaces serine or alanine (or occasionally, glutamic acid) and introduces an unexpected chymotryptic split when the haem peptide C2 is redigested with chymotrypsin after removal of the haem group. Glutamine in position 40 replaces a normally conservative serine or threonine residue. Glycine has not previously been observed in position 54, although this position exhibits some variability. Histidine occurs in position 93, replacing a residue which has been an invariant aspartic acid in all sequences published to date, with the exception of asparagine in rattlesnake cytochrome *c* (Bahl & Smith, 1965).

In the present investigation, the sequence of snail cytochrome *c* has been deduced from analyses of peptide fragments by the dansyl–Edman technique (Gray & Hartley, 1963). Peptide fragments from chymotryptic and tryptic digests gave the complete sequence with almost complete overlap. Only the chymotryptic peptides corresponding to residues 16–18 and 86–94 could not be identified. The latter peptide was located by performing a preliminary

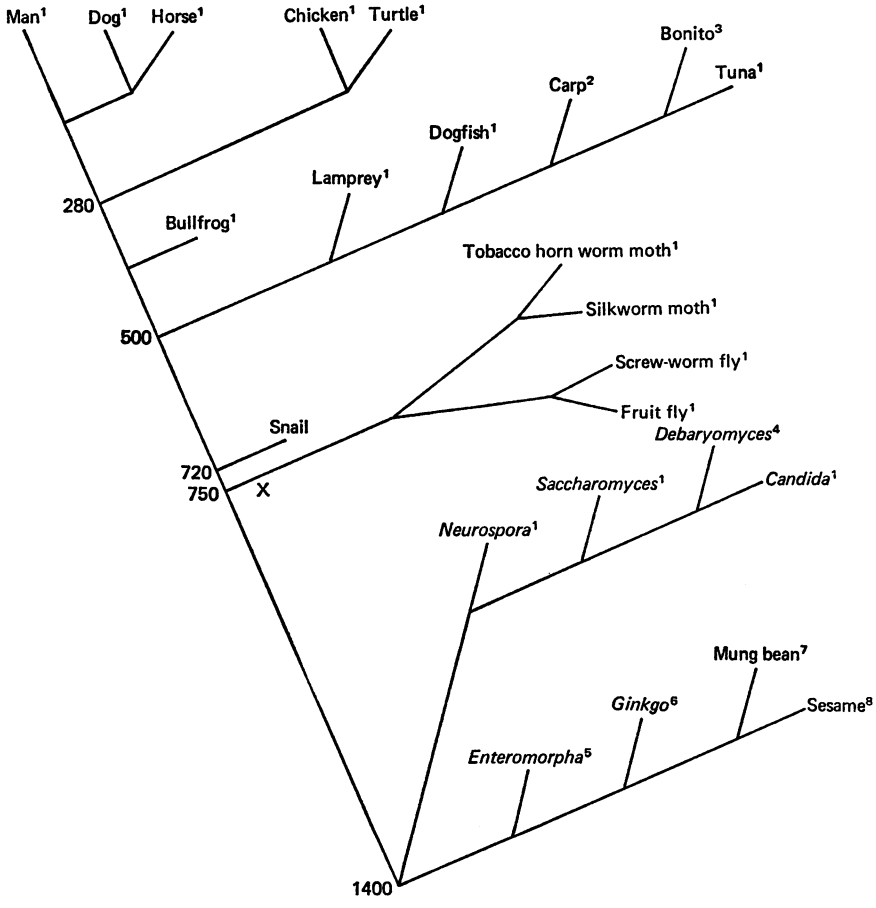

Fig. 2. *Phylogenetic relationships from cytochrome c sequence results*

Dates of divergence from the main line of descent are given in millions of years. For the meaning of X, see the Discussion section. References are as follows: 1, cited in Dayhoff (1969); 2, Gürtler & Horstmann (1970); 3, Nakayama *et al.* (1971); 4, Sugeno *et al.* (1971); 5, B. T. Meatyard, unpublished work; 6, Ramshaw *et al.* (1971); 7, Thompson *et al.* (1970); 8, Thompson *et al.* (1971).

cyanogen bromide cleavage on the cytochrome and digesting the fragment corresponding to residues 81–104 with chymotrypsin. *N*-bromosuccinimide has been used to confirm, by chemical cleavage at tryptophan in position 59, the nature of the subsequent residue. Lysine-60 was released as a new *N*-terminal amino acid after digestion.

The sequence of snail cytochrome *c* was compared with the sequences of cytochromes *c* from representatives of other major taxonomic groups. The sequences selected included all available sequences for the fishes and insects, together with representative members of other groups. The topology of the phylogenetic relationships between these species was derived by the ancestral sequence method of Dayhoff & Eck (1966). In the topology with the minimal number of amino acid substitutions (see Fig. 2) the molluscs branch from the main line of descent to the vertebrates subsequent to the divergence of the insects.

The calculation assumes that the extra *N*-terminal residues in insects, plants and fungi are significant, and each additional residue is regarded as a single amino acid difference. If the topology is calculated only over the common region of the molecule, Fig. 2 still represents the minimal topology. However, another minimum, with snail placed on the insect line, position X on Fig. 2, exists. Whether or not the *N*-terminal tails are included in the analysis, a topology with snail diverging before the insects has a greater number of amino acid substitutions than the given topology and hence is less likely than the given topology. *N*-Terminal acetyl groups have been ignored in the analysis as they do not affect the topology.

The dates of divergence of the various groups from the ancestral line have been calculated from averaged amino acid differences between groups (Margoliash & Smith, 1965), by using a 'unit evolutionary period' of 28.2 million years. The dates have not been corrected for back mutations, parallel mutations or indirect pathways (Margoliash & Smith, 1965), and hence are minimal figures. The times confirm the conclusion that the divergence of the snail (mollusc) line is close to that of the insects. It is possible, however, that further sequences for the molluscs or other invertebrates may reverse these points of divergence, but it is certain that both the molluscs and the insects diverged from the main line of descent to the vertebrates around 750 million years ago, i.e. fairly late in the Precambrian era.

We thank Dr. M. E. Boulter and Mrs. L. Mayo for invaluable assistance with the dansyl–Edman analysis, and the Nuffield Foundation and the Science Research Council for financial support.

References

Bahl, C. P. & Smith, E. L. (1965) *J. Biol. Chem.* **240**, 3583–3593

Dayhoff, M. O. (1969) *Atlas of Protein Sequence and Structure*, vol. 4, National Biomedical Research Foundation, Silver Spring, Md.

Dayhoff, M. O. & Eck, R. V. (1966) *Atlas of Protein Sequence and Structure*, vol. 2, National Biomedical Research Foundation, Silver Spring, Md.

Gray, W. R. & Hartley, B. S. (1963) *Biochem. J.* **89**, 379–380

Gürtler, L. & Horstmann, H. J. (1970) *Eur. J. Biochem.* **12**, 48–57

Margoliash, E. & Smith, E. L. (1965) in *Evolving Genes and Proteins* (Bryson, V. & Vogel, H. J., eds.), pp. 221–242, Academic Press, New York

Margoliash, E., Smith, E. L., Kreil, G. & Tuppy, H. (1961) *Nature (London)* **192**, 1121–1127

Nakayama, T., Titani, K. & Narita, K. (1971) *J. Biochem. (Tokyo)* **70**, 311–326

Ramshaw, J. A. M., Richardson, M. & Boulter, D. (1971) *Eur. J. Biochem.* **23**, 475–483

Richardson, M., Richardson, D., Ramshaw, J. A. M., Thompson, E. W. & Boulter, D. (1971) *J. Biochem. (Tokyo)* **69**, 811–813

Sugeno, K., Narita, K. & Titani, K. (1971) *J. Biochem. (Tokyo)* **70**, 659–682

Thompson, E. W., Laycock, M. V., Ramshaw, J. A. M. & Boulter, D. (1970) *Biochem. J.* **117**, 183–192

Thompson, E. W., Richardson, M. & Boulter, D. (1971) *Biochem. J.* **121**, 439–446