The primary structure of cytochrome *c* from the nematode *Caenorhabditis elegans*

Jacques R. VANFLETEREN,* Evert A. I. M. EVERS,† Gerrit VAN DE WERKEN‡ and Jozef J. VAN BEEUMEN†

*Laboratory of Animal Morphology and Systematics and †Laboratory of Microbiology and Microbial Genetics, State University of Ghent, Ledeganckstraat 35, B-9000 Ghent, Belgium, and ‡National Institute of Public Health and Environmental Protection, P.O. Box 1, 3720 BA Bilthoven, The Netherlands

The complete amino acid sequence of cytochrome *c* from the nematode *Caenorhabditis elegans* was determined. The native protein displays the same spectral properties in the oxidized and reduced states as horse heart cytochrome *c*. The apoprotein consists of 110 amino acid residues and differs from human cytochrome *c* by 44 substitutions, one internal deletion, five *N*-terminal additions and two *C*-terminal additions. One of the substitutions is the replacement of an 'invariant' phenylalanine residue at position 15 by tyrosine. The *N*-terminal sequence extension contains a short peptide motif, which is highly homologous with a peptide fragment present at the *N*-terminus of annelid and insect cytochrome *c* sequences. From the number of amino acid changes and the evolutionary rate of cytochrome *c* it would appear that nematodes diverged from a line leading to man about 1.4 billion years ago. When similar data based on the amino acid sequences of the histones H1, H2A, H2B and H3 are taken into account, the average estimate is 1.1 ± 0.1 billion years.

INTRODUCTION

It is well recognized that evolving proteins behave like 'molecular clocks' and that they can be used to construct phylogenetic trees. When calibrated properly, e.g. by means of accurately dated fossil records, they can be used to estimate divergence times between various taxa, even where fossil evidence is not available (Margoliash, 1963; Dickerson, 1971; Wilson *et al.*, 1977).

Mitochondrial cytochrome *c* is particularly well suited for the study of the phylogenetic relationships between distantly related phyla or major sections within phyla. This relatively small redox protein of around 12 kDa is known to evolve very slowly owing to the high number of functional constraints specifying the operative molecule (Dickerson, 1971). Cytochrome *c* can be purified with relative ease and can readily be sequenced by means of modern instrumentation.

Probably because of problems of obtaining adequate amounts of tissue, the most phylogenetically rewarding taxa, commonly called the lower metazoans, have not hitherto been studied by using this approach. Yet the interrelations between these taxa, the divergence of the protostome and deuterostome lines of evolution, the origin of the chordates etc. are all long-standing questions in biology that may be clarified through the study of appropriate molecular clocks.

In the present paper we give the complete primary structure of cytochrome *c* of the free-living nematode species *Caenorhabditis elegans*. From the number of amino acid substitutions observed we infer that the Phylum Nematoda and the line eventually leading to vertebrates and man diverged from each other over one billion years ago, close to the branching of animals and plants.

MATERIALS AND METHODS

Nematode culture

The wild-type strain (N2) of *C. elegans* var. Bristol was used in this study. The culture conditions have been described at length previously (Vanfleteren *et al.*, 1990). Washed nematodes were

pelleted at 2000 *g* for 10 min and suspended in an equal volume of homogenizing buffer (0.25 M-sucrose/5 mM-CaCl₂/50 mM-Tris/HCl buffer, pH 7.4). The resultant slurry was dripped into liquid N₂ and stored at –80 °C. The accumulated yield of fresh nematode tissue was approx. 2 kg.

Preparation of cytochrome *c*

The procedure used for the isolation of mitochondrial cytochrome *c* was rather unconventional because the protein was to be recovered from combined supernatant portions that were obtained when nuclei were prepared from nematode tissue for the further extraction of histones (Vanfleteren *et al.*, 1990). Briefly, 500 g portions of nematode beads were crushed in the 3.8-litre (1-U.S.-gallon) container of a Waring Blendor until a fine powder was obtained. Immediately after thawing, phenylmethanesulphonyl fluoride was added at 1 mM final concentration and the mixture was blended twice at full speed in the 1-litre cup of the apparatus, each for 2 min. The homogenate was diluted with 4 vol. of homogenizing buffer; phenylmethanesulphonyl fluoride was added at 1 mM and 2-mercaptoethanol at 1% (v/v). Intact mitochondria, nuclei and large debris were sedimented at 4000 *g* for 10 min. The pellets were suspended again in 1200 ml of homogenizing buffer containing 0.5% (v/v) Nonidet P-40, and a fresh amount of phenylmethanesulphonyl fluoride was added at 0.1 mM final concentration. Large debris was sedimented at 40 *g* for 10 min and re-extracted twice with homogenizing buffer. The three supernatant portions were combined and the nuclei were pelleted at 4000 *g* for 15 min. All supernatants (approx. 20 litres) from the last centrifugation step, which contained free cytochrome *c*, were immediately frozen and kept at –80 °C until use.

Further purification of cytochrome *c* was achieved as follows. Portions of the solution containing cytochrome *c* were thawed as needed and clarified by centrifugation at 80000 *g* for 30 min. The cytochrome was trace-enriched on a 5 cm × 25 cm column of CM-32 CM-cellulose (Whatman, Maidstone, Kent, U.K.) that had been equilibrated against 20 mM-sodium phosphate buffer, pH 7.6. A red band of cytochrome *c* formed and migrated

† To whom reprint requests should be addressed.

steadily towards the bottom of the column as more sample volume passed through the column. The chromatographic support was finally pressed out of the column with N_2 and a portion of the column support containing the red band was excised. The resin was stirred with twice-distilled water and allowed to settle again in a 2.5 cm $\times$ 40 cm column. The packed resin was washed with 20 mM-sodium phosphate buffer, pH 7.6. Cytochrome *c* was eluted with 0.5 M-NaCl in the sodium phosphate buffer. The red eluates were stored at -80°C until the entire sample was processed by this technique.

Final purification was achieved by salt-gradient ion-exchange chromatography, essentially as described by Margoliash & Walasek (1967). After being thawed, the crude preparations were dialysed for 36 h against 20 vol. of 10 mM-sodium phosphate buffer, pH 7.6, with Spectrapor (Spectrum, Los Angeles, CA, U.S.A.) dialysis tubing with a 1000 Da cut-off. All cytochrome *c* was converted into the ferric form by the addition of a few crystals of $K_3Fe(CN)_6$ and subsequently chromatographed on a column (2.5 cm $\times$ 75 cm) of CM-32 CM-cellulose under a linear gradient of 0–0.5 M-NaCl in the sodium phosphate buffer. Cytochrome *c* from *C. elegans* was eluted at 0.17–0.18 M-NaCl as determined by refractometric analysis of the peak eluate. The preparation contained approx. 12 mg of protein. It was stored at -80°C .

Warming the preparations above 4°C was carefully avoided at all stages of the procedure.

Electrophoresis

The purity of the preparation was evaluated by SDS/Tricine/PAGE (Schägger & Von Jagow, 1987). Separating gels were 14 cm wide, 12 cm high and 0.75 mm thick and contained 16.5% T and 6% C (where C is the percentage concentration of bisacrylamide relative to T, the total concentration of acrylamide and bisacrylamide). The separating gel was overlaid by a 1.5 cm-high spacer gel that contained 10% T and 3% C. The uppermost stacking gel (approx. 1 cm high) contained 4% T and 3% C. The cathodic (upper) electrolyte was 0.1% SDS/0.1 M-Tris/0.1 M-Tricine buffer, pH 8.25; the lower electrolyte was 0.2 M-Tris/HCl buffer, pH 8.9. Samples were dissolved in Laemmli (1970) sample buffer. The gels were run at 80 V until the samples had entered the stacking gel portion, and then at 120 V overnight.

Removal of haem

Two samples each containing approx. 100 nmol (1.2 mg) of cytochrome *c* were transferred to 15 ml conical polypropylene centrifuge tubes (Falcon; Becton-Dickinson, Lincoln Park, NJ, U.S.A.), and 0.1 vol. of 1 M-HCl was added. Cytochrome *c* was then quantitatively precipitated by the addition of an equal volume of a saturated solution of Reinecke salt (ammonium tetrathiocyanodiammonochromate) in 0.1 M-HCl (Lindh & Brantmark, 1965). The precipitate was sedimented at 2000 g for 10 min. Reinecke salt was extracted twice with acetone containing 0.06 M-HCl. Haem was removed from the native protein by using $HgCl_2$ in acidic urea (Ambler, 1963). The apoprotein was then recovered by precipitation with 20 vol. of 20% (w/v) trichloroacetic acid and centrifugation at 2000 g for 10 min. The pellets were washed once with acetone containing 0.06 M-HCl and twice with ethyl acetate, and finally dried under reduced pressure.

Modification of cysteine residues

The dry protein pellets were dissolved in 200 μl of 8 M-urea/50 mM-Tris/HCl buffer, pH 7.4. The tubes were flushed with a stream of N_2 . Subsequently 0.5 μl of vinylpyridine (5 μmol) was added to each sample. After standing for 2 h at room temperature, 20 vol. of 20% (w/v) trichloroacetic acid was added. The precipitated protein was pelleted at 2000 g and

washed with acetone/0.06 M-HCl, once more with acetone only and then with ethyl acetate, and finally dried under reduced pressure.

Production and separation of peptides

One sample (100 nmol of cytochrome *c*) was dissolved in 500 μl of 50 mM-acetic acid. Enough pyridine was added to raise the pH to 4.0. The protein was cleaved with *Staphylococcus aureus* V8 proteinase (EC 3.4.21.19; Miles Laboratories, Stoke Poges, Bucks., U.K.) for 40 h at 37°C , at an enzyme/substrate ratio of 1:30 (w/w) (Houmar & Drapeau, 1972). A second sample, also containing 100 nmol of apoprotein, was dissolved in 500 μl of 2% (v/v) formic acid and heated at 106°C for 4 h (Ingliš, 1983). Both peptide mixtures were separated by reversed-phase h.p.l.c. on a Vydac C-4 (Alltech, Deerfield, IL, U.S.A.) column (4.6 mm $\times$ 250 mm) connected to a stainless-steel LKB Produkter (Stockholm, Sweden) h.p.l.c. system. Peptide peaks were numbered SA_1 – SA_{47} and AH_1 – AH_{56} according to their increasing retention times.

Approx. 20 nmol of peptide SA_{35} was subcleaved with 0.4 nmol of endoproteinase Arg-C (EC 3.4.21.40; Boehringer, Mannheim, Germany) in 100 mM-ammonium bicarbonate buffer, pH 8.3, for 4 h at 37°C . The resulting peptides were resolved on a narrow-bore (2.1 mm $\times$ 22 cm) PTC C_{18} column (Applied Biosystems, Foster City, CA, U.S.A.) with a 130 A Microseparation System (Applied Biosystems). Peaks were numbered $SA_{35}RC_1$ – $SA_{35}RC_{25}$ according to their increasing retention times.

Peptide fractions AH_1 + AH_2 , which putatively contained a short blocked N-terminal peptide fragment, were dissolved in 0.05 M-HCl and heated at 110°C for 4 h.

C-Terminal sequence analysis

Separate samples containing approx. 2 nmol of apoprotein each were digested in 50 mM-sodium acetate buffer, pH 5.2, with either carboxypeptidase P (EC 3.4.16.1; Boehringer) or carboxypeptidase Y (EC 3.4.16.1; Boehringer) at an enzyme/substrate ratio of 1:125 (w/w). Norleucine was added as an internal standard at 30 pmol/ μl . Samples (5 μl) were taken at various time intervals and analysed following pre-column derivative formation with phenyl isothiocyanate on a 420 A Derivatizer linked to a 130 A Microseparation system (Applied Biosystems). Appropriate control experiments containing enzyme, internal standard and buffer only were run simultaneously to account for background free amino acids.

Automated sequence analysis

A pulsed-liquid model 477A Sequenator equipped with an on-line 120 phenylthiohydantoin analyser (Applied Biosystems) was used. The peptides were applied on to glass-fibre discs that had been coated with 3 mg of Polybrene and subjected to two Edman degradation cycles before application of the samples.

M.s.

Mass spectra were obtained on a JEOL HX110/HX110 tandem mass spectrometer of EBEB geometry, equipped with the DA5000 data system. Fast atom bombardment was applied for desorption/ionization of the sample, by use of the JEOL Xe atom gun operated at 3 kV. The instrument was operated at 10 kV accelerating potential. Glycerol was used as a matrix.

High-energy-collision-induced m.s./m.s. spectra were acquired by introducing He gas into a collision cell at ground potential so that the ion energy in the collision region was 10 keV. The collision gas was adjusted to a pressure that gave approximately a factor of 4 in attenuation of the parent ion in abundance at the final detector.

Fig. 1. Analysis of pure cytochrome *c* from *C. elegans* by SDS/Tricine/PAGE (16.5% T, 6% C)

Lane a, molecular-mass markers: myoglobin, 17.2 kDa; cytochrome *c* (horse), 12.3 kDa; aprotinin, 6.5 kDa; insulin B-chain, 3.5 kDa; insulin A chain, 2.5 kDa; bacitracine, 1.5 kDa. Lanes b–e, cytochrome *c* from *C. elegans*: lane b, 2 µg; lane c, 4 µg; lane d, 8 µg; lane e, 16 µg.

Fig. 2. Absorption spectra of oxidized (—) and reduced (-----) *C. elegans* cytochrome *c*

RESULTS

Characterization of the cytochrome *c* from *C. elegans*

The purity of the preparation can be evaluated from the gel pattern shown in Fig. 1. Contaminating proteins were only visible when the gel was heavily overloaded. The spectral properties of the nematode cytochrome *c* were very similar to those of horse heart mitochondria cytochrome *c* (Margoliash & Frohwirt, 1959), with maxima at 410–412 nm (γ peak) in the oxidized state and at 315 nm (δ peak), 415 nm (γ peak), 520 nm (β peak) and 550 nm (α peak) in the reduced state (Fig. 2). The absorbance ratio ($A_{550(\text{red.})}/A_{280(\text{ox.})}$) was 1.16:1, which is slightly less than the value 1.25:1 usually found (Margoliash & Walasek, 1967). This difference possibly reflects the presence of small amounts of contaminating protein.

Table 1. Final yields of h.p.l.c.-purified peptides from *C. elegans* cytochrome *c*

Yields were calculated from the sequencer chromatogram report files assuming 50% initial yield values. Both Sa and AH peptide fractions (see also Figs. 3–5) were derived from 100 nmol of protein.

Peptide	Yield (nmol)
Sa ₁₅	44.1
Sa ₁₆	48.6
Sa ₂₆	39.7
Sa ₃₅	23.0
Sa ₃₅ RC ₁₉	3.8
Sa ₄₅	46.1
AH ₁₊₂	2.3
AH ₃	21.0
AH ₁₉	3.8
AH ₂₁	3.8
AH ₃₂	5.4
AH ₃₃	24.8
AH ₅₀	5.4

Sequence determination

The complete amino acid sequence of *C. elegans* cytochrome *c* is represented in Fig. 8. Most of it was determined readily through extended runs in the Sequenator of major peptides released after cleavage of intact apoprotein by limited acid hydrolysis (Fig. 3) or digestion with *S. aureus* V8 proteinase (Fig. 4). The yields of h.p.l.c.-purified peptides are listed in Table 1.

Interestingly, the peptide fragments AH₁₉ and AH₂₁ (Fig. 3) represented an identical anterior region of the molecule. In peptide AH₂₁ two pyridylethylated cysteine residues were identified at positions 12 and 15, whereas in peptide 19 no amino acid phenylthiohydantoin derivatives were derived from molecules that had failed to react with vinylpyridine. As can be seen from Fig. 8, peptide overlaps between the Sa and AH peptides confirmed sufficiently that no intervening peptide fragment was missed, except at positions 51–52 of the polypeptide chain. A peptide (Sa₃₅RC₁₉) was then isolated from a sub-digest of peptide fragment Sa₃₅ that overlapped with peptide AH₅₀ and confirmed the alignment in that region of the molecule (Figs. 5 and 8). The cleavage of the Ser-53–Ala-54 peptide bond by *S. aureus* V8 proteinase was unexpected. However, non-specific cleavage of Ser–Xaa bonds by this enzyme was also reported by other workers (Evenberg *et al.*, 1977; Walker *et al.*, 1977).

Peptide fraction Sa₁₈ putatively contained a blocked *N*-terminal peptide fragment as it yielded no sequence information at all. Since many cytochrome *c* proteins have the *N*-terminal sequence *N*-acetyl-Gly-Asp- we decided to hydrolyse the combined peptide fractions AH₁ + AH₂ in the presence of 0.05 M-HCl in order to achieve partial cleavage of the putative *N*-acetyl bond. Subsequent sequencing of this material identified a major peptide Ser-Asp and a minor one Glu-Arg-Ala-Asp, which confirmed the alignment as well as the identity of the residues in the region 94–97 of the sequence.

The *N*-terminal structure was then further defined by m.s. analysis of peptide Sa₁₈. Fig. 6 shows the collision-induced-decomposition m.s./m.s. spectrum of the (*M* + *H*)⁺ ion of the peptide investigated. The amino acid sequence can be inferred from the fragmentation pattern as being acetyl-Ser-Asp-(Ile/Leu)-Pro-Ala-Gly-Asp-Tyr-Glu. The interpretation of the diagnostic fragment ions, indicated in the Roepstorff–Fohlmann nomenclature for bond identification (Roepstorff & Fohlmann,

Fig. 3. Reverse-phase h.p.l.c. of the peptides generated from *C. elegans* cytochrome *c* by limited acid hydrolysis

A peptide mixture representing approx. 100 nmol of protein was loaded on a Vydac C-4 column (4.6 mm $\times$ 250 mm) and aq. 0.1% (v/v) trifluoroacetic acid was pumped through the column at 1 ml/min for 10 min. Then the column was developed with a linear 90 min gradient of 0–60% (v/v) acetonitrile in aq. 0.1% (v/v) trifluoroacetic acid (-----) at the same flow rate. A_{220} (—) was monitored. Peaks were numbered AH_1 – AH_{86} according to their increasing retention times.

Fig. 4. Reverse-phase h.p.l.c. of peptides produced by digestion of *C. elegans* cytochrome *c* with *S. aureus* V8 proteinase

A peptide mixture representing approx. 20 nmol of protein was resolved on the Vydac C-4 column with the following gradient of acetonitrile in aq. 0.1% trifluoroacetic acid at a flow rate of 1 ml/min: 0% for 5 min, 0–15% (v/v) in 10 min, 15–40% (v/v) in 120 min and 40–60% (v/v) in 10 min (-----). A_{220} (—) was monitored. Peaks were numbered Sa_1 – Sa_{47} according to their increasing retention times.

1984), is given in Table 2. Remarkable is the very intensive peak at m/z 651, a peak that is exhibited in the conventional fast-atom-bombardment mass spectrum (not shown here) as well. The fragment ion with m/z 651 corresponds to the sequence Pro-Ala-Gly-Asp-Tyr-Glu. The m.s./m.s. spectrum of the fragment with m/z 651 was in accordance with this sequence. Apparently the bond between proline and (isoleucine/leucine) is very labile in this peptide. This phenomenon also explains the occurrence of fragment ions resulting from the fragment with m/z 651 in the

Fig. 5. H.p.l.c. map of the peptides released after sub-cleavage of peptide Sa_{35} from *C. elegans* cytochrome *c* with endoproteinase Arg-C

The peptide mixture was resolved on a small-bore PTC C_{18} column (2.1 mm $\times$ 220 mm) with the following gradient from 0% to 100% (v/v) solvent B, where solvent A was aq. 0.1% (v/v) trifluoroacetic acid and solvent B was 70% (v/v) acetonitrile in aq. 0.1% (v/v) trifluoroacetic acid: 0–57% (v/v) in 60 min, 57–86% (v/v) in 10 min and 86–100% (v/v) in 5 min (-----). A_{220} (—) was monitored. The peaks were numbered $Sa_{35}RC_1$ – $Sa_{35}RC_{24}$ according to their increasing retention times.

Fig. 6. Collision-induced-decomposition m.s./m.s. spectrum of peptide Sa_{18} from *C. elegans* cytochrome *c*

The structure-diagnostic fragment ions are indicated in the Roepstorff-Fohlmann nomenclature for bond identification. The b series represents the charged ions remaining after loss of C-terminal neutral fragments. The y' series represents the charged ions remaining after loss of N-terminal neutral fragments. * indicates m.s./m.s. fragments from the fragment ion of m/z 651: m/z 633, 578, 504, 341 and 226. The interpretation of the diagnostic fragment ions is given in Table 2.

Fig. 7. Release of amino acid residues from *C. elegans* cytochrome *c* during digestion with carboxypeptidase P

m.s./m.s. spectrum of the fragment with m/z 1008 (indicated with an asterisk in the spectrum in Fig. 6). Clearly ions with m/z 651 were present in the collision cell, although only ions with m/z 1008 were selected in the first mass spectrometer. Because of the lability of the (Ile/Leu)-Pro peptide bond, ions of m/z 651 have been formed in all regions of the mass spectrometer before and in the collision cell. Those ions of m/z 651 that were formed after the magnet of the first mass spectrometer (B1) have reached the collision cell and were partially fragmented here, resulting in the ions discussed.

Sequence analysis of peptide AH_3 eventually showed that position 3 of the amino acid chain is occupied by isoleucine.

The C-terminal amino acid assignments were confirmed by C-terminal sequence analysis with the use of carboxypeptidases. Digestions were carried out in weak acid conditions with carboxypeptidases P and Y, since the apoprotein was not soluble at pH values above 6. Carboxypeptidase Y released only leucine, whereas carboxypeptidase P released several amino acids at the rates leucine > serine > lysine > alanine (Fig. 7). These results are in very good agreement with the C-terminal amino acid assignments determined by Edman degradation of peptide AH_{33} . It is remarkable, however, that the observed enzyme specificities were rather aberrant from what is generally accepted for these exopeptidases, e.g. carboxypeptidase Y was expected to cleave off Leu-110 and Ser-109, whereas carboxypeptidase P could be expected not to cleave off Ser-109 (Fig. 8).

Table 2. Interpretation of the diagnostic fragment ions in the collision-induced-decomposition m.s./m.s. spectrum of peptide Sa₁₈ from *C. elegans* cytochrome *c*

Abbreviations: N.P., not present; imm, immonium ion; R(I/L), side chain of I/L; R(E), side chain of E.

<i>m/z</i>	Fragment type	Assignment
1008.5		MH ⁺
990.5		MH ⁺ - H ₂ O
861.3	b8	MH ⁺ - H ₂ O - E
698.4	b7	MH ⁺ - H ₂ O - E - Y
583.3	b6	MH ⁺ - H ₂ O - E - Y - D
526.4	b5	MH ⁺ - H ₂ O - E - Y - D - G
455.3	b4	MH ⁺ - H ₂ O - E - Y - D - G - A
358.3	b3	MH ⁺ - H ₂ O - E - Y - D - G - A - P
245.3	b2	MH ⁺ - H ₂ O - E - Y - D - G - A - P - (I/L)
130.2	b1	MH ⁺ - H ₂ O - E - Y - D - G - A - P - (I/L) - D
879.3	y ⁿ 1	MH ⁺ + H - Acetyl - S
764.5	y ⁿ 2	MH ⁺ + H - Acetyl - S - D
651.3	y ⁿ 3	MH ⁺ + H - Acetyl - S - D - (I/L)
554.2	y ⁿ 4	MH ⁺ + H - Acetyl - S - D - (I/L) - P
483.2	y ⁿ 5	MH ⁺ + H - Acetyl - S - D - (I/L) - P - A
N.P.	y ⁿ 6	MH ⁺ + H - Acetyl - S - D - (I/L) - P - A - G
311.3	y ⁿ 7	MH ⁺ + H - Acetyl - S - D - (I/L) - P - A - G - D
148.1	y ⁿ 8	MH ⁺ + H - Acetyl - S - D - (I/L) - P - A - G - D - Y
963.4		MH ⁺ - CO ₂ H
951.6		MH ⁺ - R(I/L)
935.6		MH ⁺ - R(E)
70.6	imm (P)	
86.5	imm (I)	
136.2	imm (Y)	

Table 3. Dating of divergence of nematodes from a line leading to mammals

The number of changes actually seen was corrected for multiple substitutions at the same position by using the Poisson correction (Dickerson, 1971). The nematode histone data are taken from the references cited: histone H1.1 from Vanfleteren *et al.* (1988), histone H2A from Vanfleteren *et al.* (1987c), histone H2B from Vanfleteren *et al.* (1986), histone H3 from Vanfleteren *et al.* (1987b) and histone H4 from Vanfleteren *et al.* (1987a). The nematode sequences were aligned and compared with those from the species indicated within parentheses (man, rabbit or cow). Deletions, additions and insertions were taken as substitutions. Differences involving only one of two amino acids at a dimorphic locus were counted as one-half. The Unit Evolutionary Period for cytochrome *c* was taken from Dickerson (1971). The evolutionary rate estimates of histone proteins were taken from Lewin's (1980) Tables 11.2 and 11.3, but were corrected for multiple hits by using the Poisson correction. The evolutionary rate of histone H4 has an indicative value only as it is based on too few changes. For that reason histone H4 was not used to calculate the average divergence time estimate, which is given as mean $\pm$ S.E.M.

Protein	Changes per 100 residues	Changes per 100 residues corrected for multiple hits	Unit Evolutionary Period (million years)	Divergence time (million years)
Cytochrome <i>c</i>	50.00 (man)	69.3	20	1390
H1.1	72.17 (rabbit)	127.9	7.9	1010
H2A	18.90 (cow)	20.9	49.6	1040
H2B	23.60 (cow)	26.9	27.6	740
H3	2.22 (cow)	2.2	571	1260
H4	0.98 (cow)	1.0	760	760
Cytochrome <i>c</i> + H1.1 + H2A + H2B + H3				1090 $\pm$ 110

DISCUSSION

Although different at many amino acid residues, the cytochrome *c* molecule from *C. elegans* shares the typical key components of spatial structure organization and function that have been identified in vertebrate cytochrome *c*. Among these are the haem-binding residues Cys-19 and Cys-22, His-23 and Met-84, which provide the axial ligands of the haem iron, Tyr-52 and Tyr(Phe)-50, which are hydrogen-bonded to the propionic acid side chains of the haem, and the invariant sequence 74-84, which

is thought to form an important portion of the surface of the molecule interacting with cytochrome *c* reductase, with Tyr-71 and Tyr-78 probably being involved in the reduction mechanism. Interaction with the cytochrome *c* oxidase is believed to occur through the aromatic residues Phe(Tyr)-15, Tyr-101 and Phe-86 (Dickerson, 1971).

Interestingly, the phenylalanine residue at position 15 is replaced by tyrosine in the nematode cytochrome *c*. This is the first report of substitution of the invariant phenylalanine at this position in a properly aligned cytochrome *c* sequence, and it

Fig. 8. Primary structure of cytochrome *c* from *C. elegans*

Automated sequence analysis was carried out on peptides generated by limited acid hydrolysis (AH) and cleavage with *S. aureus* V8 proteinase (Sa). Peptide Sa₃₅ was sub-cleaved with endoproteinase Arg-C (Sa₃₅RC). Unambiguous identification is shown as $\Rightarrow$. Assignments based on m.s. evidence are marked as $\Rightarrow$ or $\Leftarrow$. Residues released by the carboxypeptidases P and Y are shown as $\Leftarrow$. The sequence is aligned with that from human cytochrome *c* (Matsubara & Smith, 1963); alternative residues are indicated in *italics* above the amino acid chain of the nematode cytochrome *c*.

would be interesting to know whether or not there is any effect on the interaction with the cytochrome *c* oxidase.

The deletion in the nematode sequence of a histidine residue occurring at position 26 of the human sequence (Fig. 8) has not been reported to occur in any other eukaryote sequence hitherto. This deletion is not expected to alter the three-dimensional configuration of the native protein significantly; however, being located in a region of the molecule where no helix configuration is formed and where considerable size variation (from seven residues in cytochrome *c*-555 from *Chlorobium thiosulfatophilum* up to 16 residues in *Paracoccus denitrificans c*-550) is tolerated (Salemme, 1977).

The *N*-terminal sequence extension observed in the nematode sequence contains the tripeptide motif Ile-Pro-Ala, also present at the *N*-terminus in annelid cytochrome *c* and in a modified form (Val-Pro-Ala/Gln) in insect cytochrome *c* (Lyddiatt *et al.*, 1978). It is tempting to conclude that this sequence was present in the ancestral *N*-terminal sequence of animal cytochrome *c*.

Invariant residues are found at these sites where specifications to form an operative molecule are the most stringent. Though still well within limits, some divergence can be tolerated at other sites and changes accumulate with time in the amino acid sequence. Thus about 20 million years are required for one amino acid substitution to show up in every 100 amino acid residues of mitochondrial cytochrome *c*, which is called the Unit Evolutionary Period (UEP) for that particular protein (Dickerson, 1971). By comparison with human cytochrome *c* (Fig. 8, whose amino acid residues are indicated in italics), the nematode molecule shows 44 substitutions, one deletion and seven additions, or, taking the deletions and additions for as many substitutions, 50 substitutions for every 100 amino acid residues. By using Dickerson's (1971) correction for multiple hits, i.e. considering the possibility that more than one amino acid change may have occurred at the same position over a long period, this value rises to 69.3%. It then follows that nematodes have diverged from a common ancestor some 1.4 billion years ago.

Since we have sequenced all histones of *C. elegans* (Vanfleteren *et al.*, 1986, 1987*a,b,c*, 1988), it was also possible to estimate the divergence time of the Nematoda from histone evolution. The results are outlined in Table 3. All these estimates agree quite well with each other, and the inevitable conclusion is that nematodes form an ancient group that extends back to about 1.1 ± 0.1 billion years ago; this is quite soon after the divergence of plants and animals (current estimates are 1–1.2 billion years). This hypothesis also agrees well with the occurrence of an operative glyoxylate cycle in early stages of nematode development (Bolla, 1980). The glyoxylate cycle is not known to be operational in organisms other than plants and micro-organisms. Presumably this pathway was operative in the ancestors of plants and animals, but was later lost, at least partially, in the lines leading to the vertebrates and the major invertebrate phyla. In addition, nematodes are capable of synthesizing polyunsaturated fatty acids *de novo*, as are plants and single-celled eukaryotic organisms (Rothstein, 1970). So far as is known, all other animals have lost this capability.

Among the large numbers of amino acid substitutions that have accumulated over the extended period of nematode evolution, two seem to be of particular importance. The first is the substitution of tyrosine for the otherwise quite invariant phenylalanine residue at position 15, as already mentioned. The other

is the unique deletion of His-26 (human sequence) in *C. elegans* (Fig. 8). Since deletion of an amino acid involves deletion of three nucleotides while keeping the following codons in register, such an event seems much more difficult to realize than simple replacement of a single amino acid residue by another. It will therefore be interesting to see whether these mutations are also found in other nematodes. If it turns out that these changes are characteristic of nematodes as such, then a search among representatives of the lower metazoan phyla might provide new and exciting information regarding their phylogenetic interrelations. These phyla, i.e. the Porifera, the Cnidaria and Ctenophora, Rotifera, Kamptozoa, Bryozoa, Platyhelminthes and Nemertini, presumably are all 'very old', but few fossil records are available and their inter-relations remain a mystery. Further study of slowly evolving proteins in these groups will certainly prove phylogenetically very rewarding.

This work was supported by Grant no. 3.0071.84 from the Fund for Medical Scientific Research. J.R.V. is Research Director with the National Fund for Scientific Research. This paper presents research results of the Belgian national incentive programme on fundamental research in life sciences initiated by the Belgian State-Prime Minister's Office-Science Policy Programming Department.

REFERENCES

- Ambler, R. P. (1963) *Biochem. J.* **89**, 349–378
- Bolla, R. (1980) in *Nematodes as Biological Models*, vol. 2: Aging and other Model Systems (Zuckerman, B. M., ed.), pp. 165–192, Academic Press, New York
- Dickerson, R. E. (1971) *J. Mol. Evol.* **1**, 26–45
- Evenberg, A., Meyer, H., Gaastra, W., Verheij, H. M. & de Haas, G. H. (1977) *J. Biol. Chem.* **252**, 1189–1196
- Houmard, J. & Drapeau, G. R. (1972) *Proc. Natl Acad. Sci. U.S.A.* **69**, 3506–3509
- Inglis, A. S. (1983) *Methods Enzymol.* **91**, 324–332
- Laemmli, U. K. (1970) *Nature (London)* **227**, 680–685
- Lewin, B. (1980) *Gene Expression*, vol. 2: Eucaryotic Chromosomes, pp. 312–316, John Wiley and Sons, New York
- Lindh, N. O. & Brantmark, B. L. (1965) *Anal. Biochem.* **10**, 415–420
- Lyddiatt, A., Peacock, D. & Boulter, D. (1978) *J. Mol. Evol.* **11**, 35–45
- Margoliash, E. (1963) *Proc. Natl. Acad. Sci. U.S.A.* **50**, 672–679
- Margoliash, E. & Frohwirt, N. (1959) *Biochem. J.* **71**, 570–572
- Margoliash, E. & Walasek, O. F. (1967) *Methods Enzymol.* **10**, 339–348
- Matsubara, H. & Smith, E. L. (1963) *J. Biol. Chem.* **238**, 2732–2753
- Roepstorff, P. & Fohlmann, J. (1984) *Biomed. Environ. Mass Spectrom.* **11**, 601
- Rothstein, M. (1970) *Int. J. Biochem.* **1**, 422–428
- Salemme, F. R. (1977) *Annu. Rev. Biochem.* **46**, 299–329
- Schägger, H. & Von Jagow, G. (1987) *Anal. Biochem.* **166**, 368–379
- Vanfleteren, J. R., Van Bun, S. M., Delcambe, L. L. & Van Beeumen, J. J. (1986) *Biochem. J.* **235**, 769–773
- Vanfleteren, J. R., Van Bun, S. M. & Van Beeumen, J. J. (1987*a*) *Comp. Biochem. Physiol. B* **87**, 847–849
- Vanfleteren, J. R., Van Bun, S. M. & Van Beeumen, J. J. (1987*b*) *FEBS Lett.* **211**, 59–63
- Vanfleteren, J. R., Van Bun, S. M. & Van Beeumen, J. J. (1987*c*) *Biochem. J.* **243**, 297–300
- Vanfleteren, J. R., Van Bun, S. M. & Van Beeumen, J. J. (1988) *Biochem. J.* **255**, 647–652
- Vanfleteren, J. R., Van Bun, S. M., De Baere, I. & Van Beeumen, J. J. (1990) *Biochem. J.* **265**, 739–746
- Walker, J. M., Hastings, J. R. B. & Johns, E. W. (1977) *Eur. J. Biochem.* **76**, 461–468
- Wilson, A. C., Carlson, S. S. & White, T. J. (1977) *Annu. Rev. Biochem.* **46**, 573–639