Biochem. J. (1991) 274, 825–831 (Printed in Great Britain)

Rattlesnake cytochrome *c*

A re-appraisal of the reported amino acid sequence

R. P. AMBLER and M. DANIEL
Department of Molecular Biology, University of Edinburgh, Mayfield Road, Edinburgh EH9 3JR, Scotland, U.K.

The amino acid sequence of rattlesnake cytochrome *c* was originally reported in 1965, and was one of the earlier sequences to be studied. When compared with other mitochondrial cytochromes *c*, the snake sequence was soon seen to be anomalous. There were several positions in which the snake protein resembled human cytochrome *c*, although comparable anomalies were not reported for the protein from other reptiles such as lizard and turtle. Explanations of these results have included accelerated evolution in the snake lineage, paralogy rather than orthology, and faulty determination of the sequence, and the rattlesnake is now often omitted from cytochrome *c* phylogenetic trees. We have re-investigated the sequence of the snake protein, and believe that the correct sequence differs in nine places from that used for evolutionary theorizing since 1965. Four of these differences are near the haem-attachment site, in a region that was only analysed for amino acid composition in the original investigation. The other five differences are towards the *C*-terminus of the molecule, and can be explained as being due to the wrong ordering of amino acids within peptides that had been satisfactorily purified. Despite these corrections, the rattlesnake cytochrome *c* sequence still more closely resembles human cytochrome *c* than it does that of any other protein we know. We believe that this is an example of convergent evolution, although it does appear that there has been accelerated change in the line connecting the rattlesnake to the ancestral vertebrate line. Detailed evidence for the amino acid sequence of the protein has been deposited as Supplementary Publication SUP 50162 (16 pages) at the British Library Document Supply Centre, Boston Spa, Wetherby, West Yorkshire LS23 7BQ, U.K., from whom copies can be obtained on the terms indicated in Biochem. J. (1991) **273**, 5.

INTRODUCTION

Crick (1958) said "Biologists should realize that before long we shall have a subject which might be called 'protein taxonomy' – the study of the amino acid sequences of the proteins of an organism and the comparison of them between species". This prediction did come about, and the inter-species comparison of mitochondrial cytochromes *c* became the test system for the subject. Cytochrome *c* has the advantage of being almost universally distributed in eukaryotes, and in being a small and readily purified protein, often being at high concentrations. The cytochrome from horse heart was one of the first proteins to be sequenced (Margoliash *et al.*, 1961), and cytochrome *c* was the first protein for which a useful range of sequences from different species of organisms was determined (Margoliash, 1963). In the first edition of the Atlas of Protein Sequence and Structure (Dayhoff *et al.*, 1965) ten eukaryotic cytochrome *c* sequences were included, including that from rattlesnake (Bahl & Smith, 1965). Since then, the only further reptile cytochrome *c* sequences that have been reported are those from the snapping turtle *Chelydra serpentina* (Chan *et al.*, 1966) and the monitor lizard *Varanus varanus* (M. Goodman, personal communication).

Fitch & Margoliash (1967) and Dayhoff & Eck (1968) produced phylogenetic trees derived solely from cytochrome *c* sequence information, and showed that in general they agreed well with trees derived from fossil and morphological evidence. However, there were some anomalies, and those associated with the rattlesnake sequence seemed to be particularly great. Bahl & Smith (1965) had noted that there were several positions where the snake protein resembled the human one (Matsubara & Smith, 1963), and Fitch & Margoliash (1967) remarked that the turtle sequence grouped with the birds rather than with the snake. The tree in the third edition of the Atlas (Dayhoff & Eck, 1968) showed the rattlesnake on a long branch from its 'biological' position on the bird–turtle node, and the authors commented on the apparent parallel evolution with primates, but in the fourth edition (Dayhoff, 1969) the rattlesnake was no longer shown. Jukes & Holmquist (1972) interpreted the sequences as showing that evolutionary divergence in rattlesnake cytochrome *c* from the reptile stem has been more rapid than in the turtle line. Fitch (1973) considered the possibility that the published sequence might be wrong, and discounted the arguments and methods of analysis of Gibbs & McIntyre (1970) that strongly linked the snake and primate sequences. Cronquist (1976) discussed the theoretical problems involved in the taxonomic use of amino acid sequence data, and paid particular attention to the rattlesnake anomaly, and Wilson *et al.* (1977) suggested the anomaly might be explained by paralogy.

Many published amino acid sequences have been shown to be incorrect [e.g. the tentative sequence of papain (Light *et al.*, 1964) was shown to be incorrect by Drenth *et al.* (1968), and corrected by Husain & Lowe (1970) and Mitchel *et al.* (1970); *Paracoccus* cytochrome *c*-550 (Timkovich *et al.*, 1976), corrected by Ambler *et al.* (1981)], and sequencing at the DNA level is now finding many more mistakes [e.g. the *lac* repressor (Farabaugh, 1978)]. We therefore decided to re-examine the sequence of rattlesnake cytochrome *c* and to see if it really was anomalous. We have isolated cytochrome *c* from the hearts of two species of rattlesnake, and produce evidence that the sequence differs from that published by Bahl & Smith (1965) in five places. Bahl & Smith (1965) only determined an amino acid composition for the chymotryptic peptide covering the haem-attachment site, but arranged the residues in this region by comparison with the sequences of other known cytochromes *c*. We find that there are four positions in this region where the sequence that has been used for evolutionary comparisons for the last 20 years differ from that which we believe to be correct. A preliminary account of our results has been published (Ambler, 1985).

EXPERIMENTAL

Materials

Hearts of *Crotalus viridis* (western rattlesnake) were obtained from the Montana Rattlesnake Co., Harlowtown, MT, U.S.A., and of *Crotalus atrox* (western diamond-backed rattlesnake) from Dr. Michael Herron, Texas A&M College of Veterinary Medicine, College Station, TX, U.S.A. The hearts were shipped frozen to Scotland.

Preparation of rattlesnake cytochrome *c*

The procedure of Margoliash & Walasek (1967) was used. The hearts were chopped and blended for 1 min with 3 vol. of 0.3% (w/v) $Al_2(SO_4)_3,17H_2O$, and the pH was adjusted to 4.0 with 1 M-NaOH. The suspension was stirred for 2 h at room temperature (23 °C), cooled to 4 °C and filtered through muslin to remove the coarse material. The extract was then adjusted to pH 8.4 with 2 M-NH_3 solution, 6 g of Celite was added, and the mixture was then filtered through paper. The coloured solution was then passed through a column (10 cm × 2.5 cm diam.) of Amberlite CG-50 equilibrated with 0.02 M-sodium phosphate buffer, pH 8.0, and haem proteins were adsorbed as a salmon-pink band at the top of the column. The column was allowed to run dry, the resin was extruded and the coloured portion was removed by dissection. It was suspended in more of the phosphate buffer and poured into a 1 cm-diam. column, and the coloured material was carefully eluted with as small a volume as possible of phosphate buffer containing 0.5 M-NaCl. The sample was made 50% saturated with $(NH_4)_2SO_4$, a pale precipitate was removed by centrifugation, and the bulk of the colour was precipitated by 75% saturation with $(NH_4)_2SO_4$. The pellet was dissolved in 1 ml of 0.05 M-sodium acetate buffer, pH 5.1, and subjected to gel filtration through a column (90 cm × 1.5 cm diam.) of Sephadex G-50 (superfine grade), and the single protein component eluted was cytochrome *c*.

Amino acid sequence determination

The amino acid sequence was redetermined largely by methods that have been described before (Ambler & Wynn, 1973; Ambler *et al.*, 1979*a*, 1984), although the quantity of material available was at the lower limit necessary for such methods. Most of the structure has also been determined by automatic high-sensitivity sequence analysis of CNBr-cleavage fragments with an Applied Biosystems model 477A instrument.

The criteria for satisfactory results and the nature and the format of Supplementary Publication SUP 50162 are given in previous papers (Ambler & Wynn, 1973; Ambler *et al.*, 1979*a*). The results of the automatic-sequencer degradations are given in Supplementary Publication SUP 50162 in Tables of amounts of phenylthiohydantoin derivative of each amino acid recovered at each cycle.

RESULTS

Purification and properties of the protein preparations

The yield of cytochrome *c* from 350 *C. viridis* hearts (330 g) was 1.3 μmol, and that from 440 g of *C. atrox* hearts was 1.1 μmol. A batch of hearts that had thawed in transit yielded only about 15% as much cytochrome as this, and no cytochrome *c* was obtained from a batch of freeze-dried snake hearts extracted by the same method. The yield (μmol/g) was comparable with that which we have obtained from other vertebrate hearts (e.g. lion and cat).

The amino acid compositions (Table 1) and the u.v.–visible-absorption spectra of the preparations were consistent with them being pure cytochrome *c*. The $A_{280}/A_{409(ox.)}$ purity index for the preparations was 0.26, a value consistent with a cytochrome *c* containing only a singly tryptophan residue (cf. 0.22 for the horse heart protein; Margoliash & Frohwirt, 1959), and no unexpected peptides were encountered during the sequence-determination experiments. The amino acid composition was not distinguishable from that reported by Bahl & Smith (1965) for the protein from *Crotalus adamanteus* (Table 1).

Amino acid sequence

The evidence for the proposed amino acid sequence is summarized in Fig. 1. Details of the purification, analysis and sequence-determination experiments on the peptides are given in Supplementary Publication SUP 50162.

Peptides were characterized from two digests of cytochrome *c*, from which the haem had been removed before enzymic treatment. These were a tryptic digest of 0.8 μmol of *C. viridis* protein and a thermolysin digest of 1.05 μmol of *C. atrox* protein.

The large peptides from a CNBr digest of 0.12 μmol of haem-free *C. viridis* protein were separated by gel filtration through a

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

The results for *C. adamanteus* (1) are taken from Bahl & Smith (1965), calculated on the basis that Arg = 2, and derived from several separate hydrolyses. Those for *C. viridis* and *C. atrox* are from the present investigation, calculated on the basis that Σ(Ala + Leu + Asx + Glx + Phe + Met + Arg) = 37. Analysis (2) was derived from four separate hydrolyses under different conditions chosen to optimize recovery of different amino acids, and analysis (3) was from two hydrolyses. The samples for analysis (2) were: (*a*) native protein, 24 h 6 M-HCl hydrolysis; (*b*) performic acid-oxidized haem-free protein, 24 h 6 M-HCl hydrolysis (but cysteic acid peak was lost); (*c*) haem-free protein, 96 h 6 M-HCl hydrolysis; (*d*) native protein, 96 h 3 M-mercaptoethanesulphonic acid hydrolysis. The samples for analysis (3) were: (*a*) native protein, 6 M-HCl hydrolysis; (*b*) performic acid-oxidized haem-free protein, 24 h 6 M-HCl hydrolysis. All hydrolyses were performed at 105 °C. Abbreviation: N.D., not determined.

	Amino acid composition (mol of residue/mol of protein)			
	Analyses			
Amino acid	*C. adamanteus* (1)	*C. viridis* (2)	*C. atrox* (3)	Sequence (Fig. 1)
Gly	13.4	13.7	14.5	14
Ala	5.8	6.3	6.3	6
Val	3.9	4.5	(3.9)	4
Leu	6.6	6.6	6.9	7
Ile	4.5	4.8	(4.0)	5
Ser	2.9	3.8	3.6	3
Thr	9.7	9.0	9.2	10
Asp	7.8	7.7	7.9	4
Asn				4
Glu	7.7	8.4	8.6	7
Gln				1
Phe	2.8	3.1	2.8	3
Tyr	4.0	4.7	4.7	5
Trp	(1)	0.6		1
Cys	1.3	N.D.	2.0	2
Met	2.5	2.7	3.0	3
Pro	2.8	2.5	2.6	3
Lys	17.7	16.9	17.3	17
His	3.0	3.1	3.0	3
Arg	=2.0	2.0	1.8	2
				104

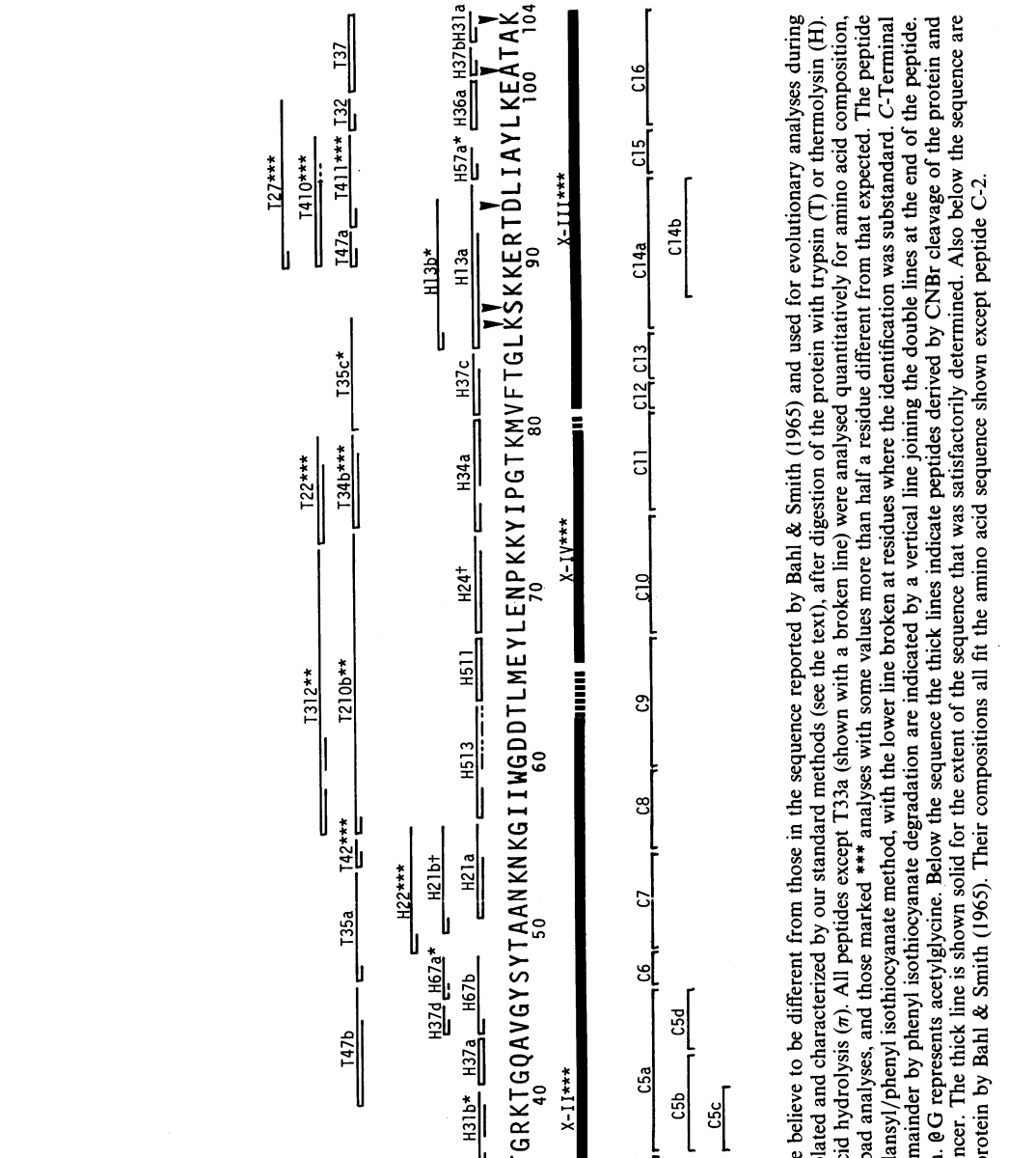

Fig. 1. Amino acid sequence of *Crotalus* cytochromes *c*

The arrowheads mark the nine positions in the sequence that we believe to be different from those in the sequence reported by Bahl & Smith (1965) and used for evolutionary analyses during the last 25 years. The peptides shown above the sequence were isolated and characterized by our standard methods (see the text), after digestion of the protein with trypsin (T) or thermolysin (H). The *N*-terminal tryptic peptide was further degraded by partial acid hydrolysis (*π*). All peptides except T33a (shown with a broken line) were analysed quantitatively for amino acid composition, but those marked * gave substandard analyses, and those marked ** bad analyses, and those marked *** analyses with some values more than half a residue different from that expected. The peptide lines are doubled where the sequence has been determined by the dansyl/phenyl isothiocyanate method, with the lower line broken at residues where the identification was substandard. *C*-Terminal residues identified as free amino acids after the removal of the remainder by phenyl isothiocyanate degradation are indicated by a vertical line joining the double lines at the end of the peptide. Peptides marked † were examined by aminopeptidase M digestion. @G represents acetylglycine. Below the sequence the thick lines indicate peptides derived by CNBr cleavage of the protein and examined in an Applied Biosystems model 477A automatic sequencer. The thick line is shown solid for the extent of the sequence that was satisfactorily determined. Also below the sequence are shown the peptides reported from a chymotryptic digest of the protein by Bahl & Smith (1965). Their compositions all fit the amino acid sequence shown except peptide C-2.

column of Sephadex G-50 (superfine grade) equilibrated with 50% (v/v) formic acid and sequenced in an Applied Biosystems model 477A instrument. The findings were in complete agreement with the results that had earlier been obtained from experiments with small peptides.

The fraction that was expected to contain the small *N*-terminal CNBr-cleavage peptide X-V was treated with pseudomonad proteinase (Drapeau, 1980), which it was hoped would remove the *N*-acetylglycine residue. The digestion mixture was analysed in the sequencer. The results were interpreted as showing some of the expected sequence Asp-Val-Glu-Lys-Gly-Lys-Lys-Ile-Phe-Ser-(Hsr), together with smaller amounts of peptide X-IV and Lys-Ile-Phe-Ser-Hsr, formed by a minor cleavage by the proteinase.

DISCUSSION

Accuracy of proposed sequence

The amino acid sequence was first re-investigated by means of the classical method of enzymic digestions followed by characterization of small peptides. These experiments suggested the nine differences from the Bahl & Smith (1965) sequence discussed below, but the amount of material available meant that the quality of the experimental results was considered to be too poor for use in a controversy. Accordingly we used the residual material for a CNBr-cleavage experiment, sequencing these CNBr-cleavage peptides in a modern automatic sequencer, and obtaining further evidence by this means for all except six of the 104 residues of the protein. The amino acid compositions of these large CNBr-cleavage peptides did not meet our normal standards of agreement with the proposed sequence. They were purified from about one-tenth of the amount of starting cytochrome (0.12 μmol) that we have hitherto used for such experiments, and the analyses showed, in particular, excessive amounts of glycine and serine, amino acids that we frequently encounter as 'fall-out' contamination in our laboratory. Nevertheless the sequences determined for these peptides were unequivocal, and completely in agreement with our expectations. We believe that most of the contamination will have been from free glycine and serine, and will have been washed out from the glass-fibre disc after loading into the Applied Biosystems model 477A sequencer.

The evidence for the *N*-terminal sequence is less satisfactory, attempts to release and identify *N*-acetylglycine by pseudomonad-proteinase digestion of small peptides derived from the *N*-terminal region having been inconclusive, so we do not have any evidence that the block at the *N*-terminus of the protein is an acetyl group. The evidence that the order of residues 1 and 2 is Gly-Asp is indirect, based on the susceptibility of peptide bond 2–3 to acid hydrolysis, and to less than definitive results of the sequencer degradation of peptide X-VD described above. Were it crucial to prove the sequence of this region, it could easily be done by mass spectrometry, but this was not considered necessary for the present purposes.

Table 2. Analyses of peptides covering the haem-binding region of *Crotalus* cytochromes

The analysis for peptide C-2 is from Bahl & Smith (1965) and analyses for peptides T210/14 and H23a are from the present investigation. The values in parentheses are those deduced from the sequence. For peptide C-2, values (*a*) are the interpretation by Bahl & Smith (1965) and (*b*) are a re-interpretation from the sequence proposed in the present investigation.

	Amino acid composition (mol of residue/mol of peptide)			
	Peptide C-2			
Amino acid	(*a*)	(*b*)	T210/14	H23a
Gly	3.30 (2)	2.67 (3)	3.2 (3)	1.4 (1)
Ala				
Val	1.09 (1)	0.88 (1)	1.0 (1)	
Leu	0.40 (0)	0.32		
Ile	0.12 (0)	0.10		
Ser	1.00 (1)	0.81 (1)	0.24 (0)	0.44 (0)
Thr	2.50 (2)	2.02 (2)	1.9 (2)	2.0 (2)
Asp			0.13 (0)	
Glu	2.50 (2)	2.02 (2)	2.1 (2)	
Phe				
Tyr				
Trp				
Cys	1.50 (2)	1.21 (2)	1.8* (2)	2.1* (2)
Met	0.80 (1)	0.65 (1)		0.7† (1)
Pro				
Lys	3.00 (3)	2.42 (2)	1.0 (1)	1.1 (1)
His	1.70 (2)	1.37 (2)	0.8 (1)	0.8 (1)
Arg				
Total	(16)		(12)	(8)
Position	11–26		14–25	12–19

* As cysteic acid.
† As methionine sulphone.

Differences between the sequences reported in 1965 and 1990

The amino acid sequence shown in Fig. 1 differs from that proposed by Bahl & Smith (1965) and used in subsequent evolutionary discussions in nine places. Four of these sites are in the haem-attachment region, characterized by Bahl & Smith (1965) in a chymotryptic peptide C-2 for which they only had sufficient material to analyse for composition. It appeared to differ from the human protein only by a threonine substituting for an isoleucine residue. Our observations were that this region contained one more glycine residue and one less lysine residue that postulated by Bahl & Smith (1965), and that the sequence differed in four places. In particular, the two amino acid residues between the haem-binding cysteine residues were different, and the two other changes (Ser-11 and Glu-22) are also unique to rattlesnake among vertebrate cytochromes *c*.

We also found five further differences between our sequence and that reported by Bahl & Smith (1965), at positions 86, 87, 93, 101 and 104, all near the *C*-terminus of the molecule. These differences are compatible with the reported amino acid compositions of the chymotryptic peptides isolated by Bahl & Smith (1965).

We have found no evidence to suggest that there are any sequence differences between the proteins from *C. atrox* and *C. viridis*. We have considered the discrepancies between our sequence and that reported by Bahl & Smith (1965) for the protein from *C. adamanteus*, and conclude that they are due to errors in interpretation of the earlier experiments, and are not due to differences between rattlesnakes. They can all be explained as the result of reasonable errors made in the interpretation of sequencing experiments in 1965, when it was not easy to determine the *N*-termini of small peptides (Ambler, 1963).

Before the introduction of the method of removing thioether-linked haem with $HgCl_2$ (Ambler, 1963), it was difficult to obtain the haem-binding-site peptides from cytochromes *c* pure and in good yield. Bahl & Smith (1965) endeavoured to purify 'peptide C-2' from a chymotryptic digest of denatured haem protein by ion-exchange chromatography on a sulphonylated polystyrene resin (Dowex 50-X2), followed by paper partition chromatography and high-voltage paper electrophoresis at pH2, but only recovered it, still impure, in 3% yield, insufficient for any further

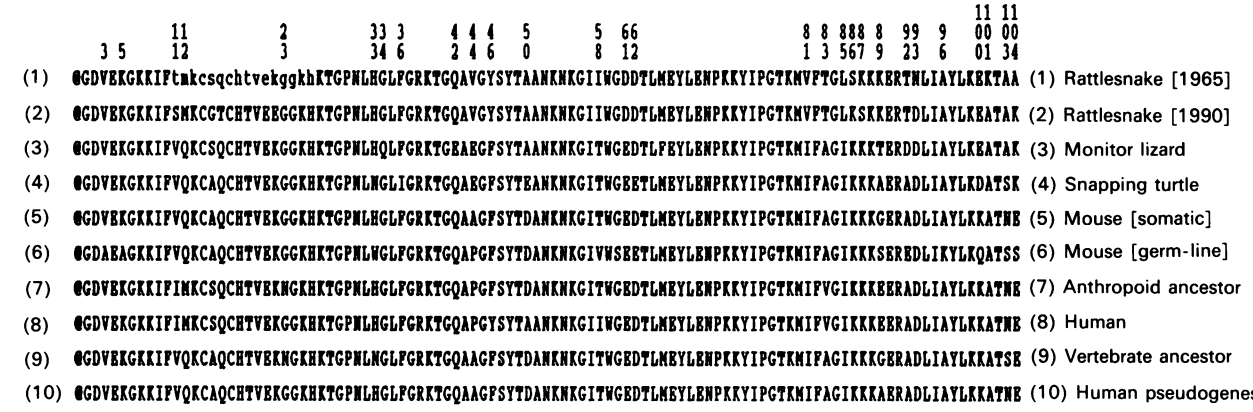

Fig. 2. Amino acid sequences of selected vertebrate cytochromes *c*

Key: (1) rattlesnake (*C. adamanteus*) (Bahl & Smith, 1965), the residues shown in lower case being aligned by comparison with other known sequences; (2) rattlesnake (*C. atrox* and *C. viridis*) (the present investigation); (3) monitor lizard (*V. varanus*) (M. Goodman, personal communication); (4) snapping turtle (*C. serpentina*) (Chan *et al.*, 1966); (5) mouse somatic cells (Hennig, 1975; Carlson *et al.*, 1977); (6) mouse germ-line (Hennig, 1975); (7) deduced anthropoid ancestor (Baba *et al.*, 1981); (8) human (Matsubara & Smith, 1963); (9) deduced vertebrate ancestor (Baba *et al.*, 1981); (10) human pseudogene consensus sequence (Evans & Scarpulla, 1988). ●G represents acetylglycine. Residues that vary as between the sequences compared are numbered.

Table 3. Difference matrix for selected vertebrate cytochromes *c*

Values are numbers of differences (out of 104) when sequences are aligned as in Fig. 2. Key: (1) rattlesnake (*C. adamanteus*) (Bahl & Smith, 1965); (2) rattlesnake (*C. atrox* and *C. viridis*) (the present investigation); (3) monitor lizard (*V. varanus*) (M. Goodman, personal communication); (4) snapping turtle (*C. serpentina*) (Chan *et al.*, 1966); (5) mouse somatic cells (Hennig, 1975; Carlson *et al.*, 1977); (6) mouse germ-line (Hennig, 1975); (7) deduced anthropoid ancester (Baba *et al.*, 1981); (8) human (Matsubara & Smith, 1963); (9) deduced vertebrate ancestor (Baba *et al.*, 1981); (10) human pseudogene consensus sequence (Evans & Scarpulla, 1988).

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
(1) Rattlesnake [1965]	–	8	17	22	19	24	18	14	21	19
(2) Rattlesnake [1990]	8	–	16	20	18	23	18	14	20	18
(3) Monitor Lizard	17	16	–	12	10	16	13	13	12	10
(4) Snapping Turtle	22	20	12	–	9	12	14	15	8	8
(5) Mouse [somatic]	19	18	10	9	–	12	7	9	3	1
(6) Mouse [germ-line]	24	23	16	12	12	–	16	17	12	12
(7) Anthropoid ancestor	18	18	13	14	7	16	–	4	8	7
(8) Human	14	14	13	15	9	17	4	–	12	9
(9) Vertebrate ancestor	21	20	12	8	3	12	8	12	–	4
(10) Human pseudogenes	19	18	10	8	1	12	7	9	4	–

sequence work. The amino acid composition evidence for peptides covering the haem-binding region is shown in Table 2. Bahl & Smith (1965) explained their excess glycine as being due to contamination. It seems possible that what we now see as excess lysine, leucine, threonine and glutamine in their peptide C-2 arose through contamination with peptide C-14b (Lys-Lys-Glu-Arg-Thr-Asp-Leu), which was in the same ion-exchange fraction, although neither alanine nor arginine was noted in the analysis. Bahl & Smith (1965) interpreted the analysis to show that the peptide differed from the corresponding human one (Matsubara & Smith, 1963) only by having *N*-terminal threonine rather than isoleucine. Our results indicate four differences between the sequence deduced by composition and homology in 1965 and that which we believe to exist (Fig. 1).

Five further differences occur in the *C*-terminal part of the sequence. Bahl & Smith (1965) do not seem to have had any evidence that their peptide C-14aT-2 was Ser-Lys-Lys rather than Lys-Ser-Lys, and hence have residues 86 and 87 transposed compared with our sequence. In the human sequence, both residues are reported to be lysine (Matsubara & Smith, 1963). Bahl & Smith (1965) deduced that residue 93 was asparagine rather than aspartic acid because of the neutral electrophoretic mobility of peptide C-14aT-6 (Thr-Asx-Leu) at pH 4.7. It is now realized that pK_a values for peptide carboxy groups are around pH 4.5, and so electrophoresis at pH 4.7 is not useful for deducing amide locations.

The evidence of Bahl & Smith (1965) for the *C*-terminus of the molecule (residues 101–104) was derived from their peptide C-16T-2. They deduced the sequence Glu-Lys-Thr-Ala-Ala from the rate of release of amino acids from the peptide by carboxypeptidase A, although the method by which the amounts of amino acid were measured was not given. At this period it was generally believed that carboxypeptidase A did not normally remove lysine residues (Fraenkel-Conrat *et al.*, 1956), and so any released lysine may have been overlooked. Subsequent experience (Ambler, 1967) showed that lysine (though not arginine) is normally released from peptides and proteins by commercial carboxypeptidase preparations.

The difficulties that Bahl & Smith (1965) encountered with sequencing small pure peptides that they had obtained in good yield emphasize the problems that confronted protein chemists at this time. The 1-fluoro-2,4-dinitrobenzene method was not easy

to use to determine *N*-termini of small peptides, as the dinitrophenol formed as a by-product of the reaction interfered with the identification of 2,4-dinitrophenyl-amino acids. The direct identification of phenylthiohydantoins was difficult by paper chromatography, and it was not until the introduction of the dansyl method (Gray & Hartley, 1963) that these difficulties were eased.

Evolutionary implications

The N.B.R.C. Protein Data Base at present records 36 vertebrate mitochondrial cytochrome *c* sequences, and 53 from other eukaryotes. In 1990, the corrected rattlesnake sequence still resembles human cytochrome *c* protein more than it does any other protein in the N.B.R.C. Protein Data base (Fig. 2 and Table 3). This similarity had been noted by Bahl & Smith (1965), using what we now believe to be an incorrect rattlesnake sequence, and, as discussed above, a variety of explanations have been put forward to account for this unexpected finding.

Hennig (1975) showed that there were tissue-specific isocytochromes *c* in vertebrates when he isolated a protein from mouse testis that differed in 13 positions from the somatic cell protein, and isocytochromes *c* have also been characterized from larval invertebrates and yeast. Scarpulla *et al.* (1982) and Evans & Scarpulla (1988) have investigated the cytochrome *c* genes and pseudogenes in man and rat, without apparently finding evidence for further tissue-specific isocytochromes. However, they have made the unexpected discovery that the majority of the many human pseudogenes would encode proteins that resemble the sequence of the functional mouse somatic protein much more closely than they resemble the human one, and are close in sequence to the deduced vertebrate ancestral sequence (Baba *et al.*, 1981; Fig. 2 and Table 3).

It is possible that the rattlesnake and the human cytochromes *c* are the products of homologous genes that are both only paralogous to the functional somatic proteins of other vertebrates, but there is no evidence (e.g. from studies on rat genes) for the existence of such a gene in other vertebrates.

In the set of vertebrate cytochromes *c* shown in Fig. 2, the only positions in which the rattlesnake protein has unique residues are 16, 81, 83, 85 (with one exception) and 87. Position 16 (at the haem-binding site) is glutamine in almost all eukaryotes, with some of the earlier reports for other residues in this position [e.g. baker's yeast (Narita & Titani, 1969) and *Neurospora crassa* (Heller & Smith, 1966)] being incorrect (Lederer & Simon, 1974), but threonine is present in this position in several bacterial cytochromes c_2 (Ambler *et al.*, 1979*b*). Position 81 is isoleucine in all other vertebrates, but is valine in rattlesnake, moths and the higher plants. Residue 83 is threonine in rattlesnake, but alanine in other vertebrates, except for the primates, where it is valine. Residue 85 is leucine in rattlesnake, dogfish, fungi and higher plants, but isoleucine in other vertebrates. Residue 87 would seem to be uniquely serine in rattlesnake, as compared with lysine in other eukaryotes, with the exception of *Euglena*, where it is alanine, and in the aberrant *Tetrahymena pyriformis*.

We believe that the sequence anomaly is best explained by there having been rapid evolution in the line leading to the Colubridae, and that some convergence of the primate and this snake line has occurred. We can think of no rationale to explain such convergence for these cytochromes *c*, as has been adduced for the stomach lysozymes of ruminants and leaf-eating primates (Stewart *et al.*, 1987), where the convergent evolution of a fermentative foregut in two groups of mammals appears to be complemented by convergence in their bacteriolytic lysozymes.

In the Superfamily Colubroidea to which the rattlesnakes (Family Viperidae) belong, the loss of features common to other reptiles such as limb girdles is more complete than in the other snake Superfamilies (Typhlopoidea and Booidea) (Underwood, 1967; Smith *et al.*, 1971). It will be interesting to determine the sequence of the cytochrome *c* from a 'primitive' (Romer, 1966) booid snake and see if this does resemble the putative common reptile line more closely than does the rattlesnake sequence.

We thank Dr. Michael Herron, Texas A&M College of Veterinary Medicine, and the Montana Rattlesnake Co. for help in obtaining rattlesnake hearts. We also thank Dr. Linda Fothergill-Gilmore, Miss Linda Kerr and Mr. Andrew Cronshaw of the WELMET Protein Characterization Facility, Department of Biochemistry, University of Edinburgh, for the automatic sequence determination, and the Wellcome Trust and the Salvesen Foundation for supporting this facility. We thank Dr. G. R. Drapeau for the gift of pseudomonad proteinase.

REFERENCES

Ambler, R. P. (1963) Biochem. J. **89**, 349–378
Ambler, R. P. (1967) Methods Enzymol. **11**, 143–154
Ambler, R. P. (1985) in Computer-Assisted Bacterial Systematics (Jones, D., Goodfellow, M. & Priest, F. G., eds.), pp. 307–335, Academic Press, London
Ambler, R. P. & Wynn, M. (1973) Biochem. J. **131**, 485–498
Ambler, R. P., Daniel, M., Meyer, T. E., Bartsch, R. G. & Kamen, M. D. (1979*a*) Biochem. J. **177**, 819–823
Ambler, R. P., Daniel, M., Hermoso, J., Meyer, T. E., Bartsch, R. G. & Kamen, M. D. (1979*b*) Nature (London) **278**, 659–660
Ambler, R. P., Meyer, T. E., Kamen, M. D., Schichman, S. A. & Sawyer, L. (1981) J. Mol. Biol. **147**, 351–356
Ambler, R. P., Daniel, M., Melis, K. & Stout, C. D. (1984) Biochem. J. **222**, 217–227
Baba, M. L., Darga, L. L., Goodman, M. & Czelusniak, J. (1981) J. Mol. Evol. **17**, 197–213
Bahl, O. P. & Smith, E. L. (1965) J. Biol. Chem. **240**, 3585–3593
Carlson, S. S., Mross, G. A., Wilson, A. C., Mead, R. T., Wolin, L. D., Bowers, S. F., Foley, N. T., Muijsers, A. O. & Margoliash, E. (1977) Biochemistry **16**, 1437–1442
Chan, S. K., Tulloss, I. & Margoliash, E. (1966) Biochemistry **5**, 2586–2597
Crick, F. H. C. (1958) Symp. Soc. Exp. Biol. **12**, 138–163
Cronquist, A. (1976) Brittonia **28**, 1–27
Dayhoff, M. O. (ed.) (1969) Atlas of Protein Sequence and Structure, vol. 4, National Biomedical Research Foundation, Silver Spring
Dayhoff, M. O. & Eck, R. V. (eds.) (1968) Atlas of Protein Sequence and Structure 1967–68, National Biomedical Research Foundation, Silver Spring
Dayhoff, M. O., Eck, R. V., Chang, M. & Sochard, M. R. (eds.) (1965) Atlas of Protein Sequence and Structure, National Biomedical Research Foundation, Silver Spring
Drapeau, G. R. (1980) J. Biol. Chem. **255**, 839–840
Drenth, J., Jansonius, J. W., Koekoek, R., Swen, H. M. & Wolthers, B. G. (1968) Nature (London) **218**, 929–932
Evans, M. J. & Scarpulla, R. C. (1988) Proc. Natl. Acad. Sci. U.S.A. **85**, 9625–9629
Farabaugh, P. J. (1978) Nature (London) **274**, 765–769
Fitch, W. M. (1973) Annu. Rev. Genet. **7**, 343–380
Fitch, W. M. & Margoliash, E. (1967) Science **155**, 279–284
Fraenkel-Conrat, H., Harris, J. I. & Levy, A. L. (1956) Methods Biochem. Anal. **2**, 359–391
Gibbs, A. J. & McIntyre, G. A. (1970) Eur. J. Biochem. **16**, 1–11
Gray, W. R. & Hartley, B. S. (1963) Biochem. J. **89**, 379–380
Heller, J. & Smith, E. L. (1966) J. Biol. Chem. **241**, 3165–3180
Hennig, B. (1975) Eur. J. Biochem. **55**, 167–183
Husain, S. S. & Lowe, G. (1970) Biochem. J. **116**, 689–692
Jukes, T. H. & Holmquist, R. (1972) Science **177**, 530–532
Lederer, F. & Simon, A.-M. (1974) Biochem. Biophys. Res. Commun. **56**, 317–323
Light, A., Frater, R., Kimmel, J. R. & Smith, E. L. (1964) Proc. Natl. Acad. Sci. U.S.A. **52**, 1276–1283
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

Margoliash, E., Smith, E. L., Kreil, G. & Tuppy, H. (1961) Nature (London) **192**, 1125–1127
Matsubara, H. & Smith, E. L. (1963) J. Biol. Chem. **238**, 2732–2753
Mitchel, R. E. J., Chaiken, I. M., & Smith, E. L. (1970) J. Biol. Chem. **245**, 3485–3492
Narita, K. & Titani, K. (1969) J. Biochem. (Tokyo) **65**, 259–267
Romer, A. S. (1967) Vertebrate Palaeontology, 3rd edn., University of Chicago Press, Chicago
Scarpulla, R. C., Agne, K. M. & Wu, R. (1982) Proc. Natl. Acad. Sci. U.S.A. **79**, 739–743
Smith, H. M., Smith, R. B. & Sawin, H. L. (1971) J. Herpetol. **11**, 115–121
Stewart, C.-B., Schilling, J. W. & Wilson, A. C. (1987) Nature (London) **330**, 401–404
Timkovich, R., Dickerson, R. E. & Margoliash, E. (1976) J. Biol. Chem. **251**, 2197–2206
Underwood, G. (1967) A Contribution to the Classification of Snakes, British Museum (Natural History), London
Wilson, A. C., Carlson, S. S. & White, T. J. (1977) Annu. Rev. Biochem. **46**, 573–639

Received 29 May 1990/23 August 1990; accepted 25 September 1990