Amino Acid Sequence of a Cytochrome *c* from the Common Pacific Lamprey, *Entosphenus tridentatus*†

Chris Nolan, Walter M. Fitch,‡ Thomas Uzzell,§ Larry J. Weiss, and Emanuel Margoliash*

ABSTRACT: Peptides were isolated from a tryptic digest of a cytochrome *c* from the common Pacific lamprey, *Entosphenus tridentatus*, and the amino acid sequences of peptides accounting for all the amino acid residues in the protein were determined. The order of the peptides in the protein was deduced by analogy with cytochromes *c* of known sequence from other eukaryotic species. The results show that the lamprey has a typical vertebrate cytochrome *c* in that it consists of a single chain of 104 residues with N-terminal *N*-acetylglycine and the characteristic clustering of hydrophobic residues and of basic residues. The unvaried metazoan sequence of residues 70–80 and the histidyl residue at position 18 are present, as well as other residues to which functional or structural roles have been assigned, such as tyrosyl residues 48, 67, and 74, tryptophanyl residue 59, phenylalanyl residue 82, and arginyl residues 38 and 91. From a statistical phylogenetic tree based on cytochrome *c*, it appears that the cyclostomes, elasmobranchs, and teleosts are more closely related to each other than to the higher vertebrates, implying the latter diverged from the former before the lower groups diverged from each other. The examination of further appropriate cytochromes *c* is needed to confirm this unexpected tentative conclusion.

The amino acid sequences of the cytochromes *c* of more than 50 eukaryotic organisms, including vertebrate, invertebrate, plant, fungal, and protozoan species, have been determined (for compilations, see Wojciech and Margoliash, 1970; Dickerson, 1972; Dayhoff, 1972). These data have provided information about structure–function relationships in cytochrome *c*, the evolutionary behavior of proteins in general, and the parallel between the degree of structural relatedness among proteins of an orthologous set[1] and the phylogenetic relationships of the host species (Fitch and Margoliash, 1967; Nolan and Margoliash, 1968; Margoliash *et al.*, 1968; Fitch and Margoliash, 1970; Dickerson, 1971; Margoliash, 1972).

Although the amino acid sequences of the cytochromes *c* of a rather large variety of mammalian species are known, relatively few cytochromes from lower vertebrates and none from species of the class cyclostomata have been characterized. This together with the fact that the lamprey, as a cyclostome, is among the most primitive of the living vertebrates prompted the present study.

Materials and Methods

Enzymes. Trypsin and chymotrypsin were crystalline preparations from Worthington Biochemical Corp. The trypsin was treated with Tos-PheCh$_2$Cl[2] (Kostka and Carpenter, 1964) before use.

Cytochrome c. Pacific lampreys, *Entosphenus tridentatus*,[3] were collected from the Willamette River near Oregon Falls in Oregon and supplied frozen by the Wilbur Ellis Co. Cytochrome *c* was extracted and purified from homogenates of the whole animals by the procedure of Margoliash and Walasek (1967), except that in step 2 the filtrate from step 1 was adjusted to pH 8.8 rather than pH 8.2–8.5 prior to filtration. The yield was 27 mg of cytochrome *c*/kg of body weight (wet), excluding three additional, minor cytochrome *c* chromatographic fractions which, together, accounted for about 6% of the total; these were not pooled with the major fraction and were not characterized. A minor protein contaminant was removed from the purified cytochrome by a final fractionation by gel filtration. Approximately 300 mg of cytochrome *c* was chromatographed on a 3.8 × 156 cm column of Sephadex G-50 fine beads (Pharmacia) in 0.1 M NH_4HCO_3. It eluted as a single, symmetrical peak following the protein contaminant. The $A_{550\ nm}$ (reduced) to $A_{280\ nm}$ (oxidized) ratio of the preparation in 0.1 M NH_4HCO_3 (pH 8.0–8.1) was 1.24.

Tryptic Digestion of Cytochrome c and Fractionation of Peptides. The cytochrome (200 mg) at a concentration of 10 mg/ml in 0.05 M NH_4HCO_3 was incubated with 2.0 mg of Tos-PheCH$_2$Cl-treated trypsin at 37°. The digestion was stopped after 4 hr by rapid freezing and lyophilization. The lyophilized digest was dissolved in 0.2 M pyridine–acetate buffer (pH 3.2) (7 ml) and a gelatinous, insoluble residue was removed by centrifugation and washed successively with buffer (four 2-ml portions) and, finally, with 6 ml of 50% (v/v) acetic acid. The supernatant fractions were combined to form the "soluble" tryptic peptide fraction; the insoluble residue is designated the "insoluble" tryptic peptide fraction.

† From the Department of Molecular Biology, Abbott Laboratories, North Chicago, Illinois 60064. *Received June 1, 1973.*

‡ On leave from the Department of Physiological Chemistry, University of Wisconsin School of Medicine, Madison, Wisconsin 53706; supported by National Science Foundation Grants GB 7486 and 144-C711.

§ Present address: The Academy of Natural Sciences, Philadelphia, Pa. 19103.

* Present address: Department of Biological Sciences, Northwestern University, Evanston, Illinois 60201.

[1] *Orthologous* proteins (Fitch and Margoliash, 1970), in addition to having descended from a common ancestral form, have been represented in that ancestor and in all the descendent species considered by single genes, so that the lineage of the genes reflects precisely, in a one-to-one fashion, the lineage of the species. Myoglobin and hemoglobin chains, for example, may well have descended from a single ancestral form (they are *homologous*), but together are not orthologous; they are termed *paralogous* and cannot be used indiscriminately to infer phylogenetic relations by comparisons of the myoglobins of some species with the hemoglobins of others.

[2] Abbreviations used are: Tos-PheCH$_2$Cl, L-1-tosylamido-2-phenylethyl chloromethyl ketone; amino acid>PhNCS, phenylthiohydantoin derivative of the amino acid indicated.

[3] We thank Miss Pearl Sonoda of the Field Museum of Natural History, Chicago, Ill., for confirming the identification of the lampreys (Field Museum of Natural History No. 73922).

"Soluble" Tryptic Peptides. The soluble peptide fraction was fractionated by chromatography on AG 50W-X2 (Bio-Rad) (Figure 1). The resin was a mixture of equal volumes of 80–200 and 250–325 wet mesh beads. Details of the chromatography are given in the legend to Figure 1. The fractions were pooled as shown in the figure and concentrated to dryness on a rotary evaporator at 40° under reduced pressure. Further purification of the peptides, where necessary, was achieved by gel filtration on Sephadex G-25 columns (1.9 × 155 cm) in 1 N acetic acid, free-flow electrophoresis on a Brinkman Model FF-1 apparatus in pyridine–acetate buffer (pH 6.5) (Nolan *et al.*, 1971) or on Whatman 3MM paper by electrophoresis at pH 6.5 in pyridine–acetate buffer or chromatography in 1-butanol–pyridine–acetic acid–water (60:40:12:48, v/v), as indicated in Table II. All purified peptides were characterized by peptide mapping as described below.

"Insoluble" Tryptic Peptides. An aliquot (0.01 ml) of the insoluble tryptic peptide fraction finely suspended in 2.5 ml of water was centrifuged, the supernatant solution was discarded, and the residue was dissolved in 0.6 ml of anhydrous trifluoroacetic acid. The solution was immediately diluted to 2.0 ml with 50% (v/v) acetic acid and chromatographed on a 0.9 × 61 cm column of Sephadex G-50 (fine beads) in 50% acetic acid. Once dissolved, the peptides remained in solution in this eluent and eluted in a single peak, detected with ninhydrin reagent after alkaline hydrolysis (Hirs *et al.*, 1956). Aliquots of this fraction, designated peptides T-12 and T-13, were analyzed for amino acids after acid hydrolysis for 24 and 96 hr (Table III). The remaining, unfractionated "insoluble" peptide fraction was digested with chymotrypsin without further fractionation. The insoluble peptides, finely suspended in 10 ml of 0.05 M NH_4HCO_3, were incubated at 37° with 1.0 mg of chymotrypsin until dissolution was complete (7 hr). The digest was fractionated on a column (0.9 × 150 cm) of AG 50W-X2 as described in the legend of Figure 2.

Heme Peptide. The heme was removed from the heme peptide (T-5) by the method of Ambler (1963), and the resulting cysteine residues were carboxymethylated as described by Crestfield *et al.* (1963).

Analytical Procedures. Amino acid analyses and Edman degradations were performed as previously described (Nolan *et al.*, 1971). For analysis of peptides, acid hydrolysis was performed for 22–24 hr unless otherwise indicated; cytochrome *c* samples were hydrolyzed for 24, 48, and 72 hr. PhNCS derivatives of amino acids were identified by thin layer chromatography on silica gel sheets with an ultraviolet indicator (Eastman Chromagram sheets, Type 6060) in solvents IV and V of Jeppson and Sjöquist (1967). The derivatives were detected on the chromatograms by uv absorption and with a ninhydrin reagent, using a modification of a method developed by F. Perini and E. Margoliash (1968, unpublished data), as follows. After examination under uv light, the chromatogram was sprayed with an 0.1% solution of ninhydrin in 95% ethanol, air-dried at room temperature, and placed in a 110° oven for 2–5 min. At this temperature, it was necessary to place the chromatogram so that the bottom surface was in contact with the metal bottom of the oven. A variety of distinctive colors are obtained as follows: Asn>PhNCS, pink to yellow-brown; Asp>PhNCS, pink; Thr>PhNCS, yellow; Ser>PhNCS, purple; Glu>PhNCS and Gln>PhNCS, olive-green; CM-Cys>PhNSC, purple; Gly>PhNCS, orange; Ala>PhNCS, purple; Met>PhNCS, yellow-brown; Leu>PhNCS, faint pink; Tyr>PhNCS, lemon yellow; Phe>PhNCS, faint yellow; Trp>PhNCS, yellow; Lys>PhNCS, faint pink; His>PhNCS, pink to brown; Arg>PhNCS, purple. Most of these are detectable with the ninhydrin reagent at levels down to about 0.005 μmol. The valine, isoleucine, and proline derivatives give no color, and those of leucine, phenylalanine, and lysine stain weakly. While this work was in progress, similar results with a ninhydrin–collidine reagent were reported by Roseau and Pantel (1969).

Peptide mapping was performed on Whatmann 3MM paper as previously described (Nolan and Margoliash, 1966); electrophoresis was in pyridine–acetate buffer at pH 6.5 (pyridine, 5; acetic acid, 0.2; water, 95; v/v) and chromatography in 1-butanol–pyridine–acetic acid–water (60:40:12:48, v/v). Peptide maps were developed with ninhydrin reagent and, in some instances, with reagents specific for tryptophan, arginine, sulfur-containing amino acids, and tyrosine and histidine, as previously described (Nolan and Margoliash, 1966).

Peptide Nomenclature. Tryptic peptides are numbered according to their occurrence in the protein sequence, T-1 being the peptide derived from the N-terminus. Chymotryptic peptides derived from tryptic peptides are similarly numbered to indicate their position in the tryptic peptide relative to the N-terminus; T-12-Cl, for example, is a chymotryptic peptide derived from the amino-terminus of T-12.

TABLE I: Amino Acid Composition of Lamprey Cytochrome *c* (Residues/Molecule).

	From Acid Hydrolysates[a]				
Residue	24 hr	48 hr	72 hr	Av or Extrapolated Value	By Sequence Anal.
Lys	18.10	17.98	17.60	17.89	18
His	2.02	2.03	1.88	1.98	2
Arg	1.84	1.87	1.96	1.89	2
Asx	7.21	7.24	7.30	7.25	7
Thr	6.87	6.75	6.18	7.2[c]	7
Ser	5.85	5.61	5.11	6.2[c]	6
Glx	11.30	11.21	11.27	11.26	11
Pro	4.18	4.06		4.12	4
Gly	12.41	12.22	11.13	11.92	12
Ala	5.01	5.00	5.20	5.07	5
Cys/2	1.33	1.63		1.63	2
Val	4.92	5.49	5.44	5.44[d]	6
Met[b]	0.91	0.73		0.82	1
Ile	3.84	4.35	4.66	4.66[d]	5
Leu	6.05	5.89	6.58	6.17	6
Tyr	3.56	3.62	3.37	3.52	4
Phe	4.67	4.75	5.33	4.92	5
Trp					1
Total					104

[a] The values given for each hydrolysis time are averages from duplicate analyses. [b] The methionine values are the sum of methionine and methionine sulfoxides. [c] Extrapolated values. [d] The 72-hr values for Val and Ile are used.

Results

The amino acid composition data for lamprey cytochrome *c* are given in Table I. The results obtained with acid hydrolysates of the protein are in good agreement with those from sequence analysis of the tryptic peptides. The somewhat low

TABLE II: Amino Acid Compositions of Soluble Tryptic Peptides.[a]

	T-1 (Residues 1–5)	T-2 (Residues 6–7)	T-3 (Residues 6–8)	T-4 (Residues 8–13)	T-5 (Residues 14–22)	T-6 (Residues 23–25)
Residue[b]						
Lys	0.99 (1)	1.00 (1)	1.96 (2)	1.83 (2)	1.10 (1)	0.96 (1)
His					1.04 (1)	
Arg						
Asx	0.96 (1)					
Thr					0.90 (1)	
Ser					0.87 (1)	0.11
Glx	1.01 (1)			1.00 (1)	2.16 (2)	
Pro						
Gly	1.02 (1)	1.00 (1)	1.02 (1)			1.13 (1)
Ala						0.92 (1)
Cys/2					1.55 (2)	
Val	1.01 (1)			1.87 (2)	1.01 (1)	
Met						
Ile						
Leu						
Tyr						
Phe				1.12 (1)		
Totals	(5)	(2)	(3)	(6)	(9)	(3)
Electrophoretic migration (cm)[d]	−8.6	+7.8	+13.4	+4.5	+0.2	+7.8
Color reactions[e]					Pauly, heme color	Gray
% yield	84	19	52	18	61	90
Purification procedure[f]				FFE, GF	FFE	

	T-7 (Residues 26 and 27)	T-8 (Residues 28–38)	T-9 (Residues 39–53)	T-9A (Residues 39–53)	T-10 (Residues 40–53)	T-11 (Residues 54–55)
Residue[b]						
Lys	0.90 (1)		1.95 (2)	2.20 (2)	1.10 (1)	1.09 (1)
His	1.14 (1)					
Arg		1.25 (1)				
Asx		1.00 (1)	2.08 (2)	2.07 (2)	2.09 (2)	
Thr		0.93 (1)	1.91 (2)	1.97 (2)	2.00 (2)	
Ser			0.95 (1)	1.10 (1)	0.91 (1)	0.91 (1)
Glx		0.95 (1)	1.00 (1)	0.97 (1)	0.94 (1)	
Pro		1.07 (1)	0.98 (1)	1.00 (1)	1.02 (1)	
Gly		3.00 (1)	2.20 (2)	2.10 (2)	2.08 (2)	
Ala			2.12 (2)	2.00 (2)	1.89 (2)	
Cys/2						
Val						0.10
Met						
Ile			0.12			
Leu		1.87 (2)				
Tyr			0.66 (1)	0.77 (1)	0.63 (1)	
Phe		0.86 (1)	1.07 (1)	0.87 (1)	1.03 (1)	
Totals	(2)	(11)	(15)	(15)	(14)	(2)
Electrophoretic migration (cm)[d]	+9.8	+2.3	+2.9	+1.9	+0.2 (N)	+7.8
Color reactions[e]	Pauly, gray	Gray, Sakaguchi	Pauly	Pauly	Pauly, yellow	Yellow
% yield	64	33	12		27	68
Purification procedure[f]		GF	PE	GF		

[a] The values in parentheses are the nearest whole number of residues. Amino acids present in amounts equivalent to <0.1 residue are not included. [b] Tryptophan was absent from the soluble tryptic peptides as determined with Ehrlich's reagent on peptide maps. [c] This value is the sum of methionine and methionine sulfoxide values. [d] Electrophoretic migration is the distance traveled during electrophoresis on paper at pH 6.5 under the conditions described under Methods. A plus sign indicates migration toward the cathode and a minus sign migration toward the anode. N denotes electrophoretic neutrality. The values are not corrected for electroendosmosis. [e] All peptides gave a blue color with ninhydrin except those described as gray, or yellow, which refers to the

TABLE II (*Continued*)

	T-14 (Residues 73–79)	T-15 (Residues 74–79)	T-16 (Residues 80–86)	T-17 (Residues 87–91)	T-18 (Residues 88–91)	T-19 (Residues 88–92)
Residue[b]						
Lys	2.15 (2)	1.03 (1)	1.00 (1)	0.97 (1)		1.06 (1)
His						
Arg				0.82 (1)	1.01 (1)	0.89 (1)
Asx		0.19				0.15
Thr	1.00 (1)	0.99 (1)				0.13
Ser	0.14	0.20		0.15		0.13
Glx	0.21	0.14		0.29 (2)	2.02 (2)	1.87 (2)
Pro	0.88 (1)	0.94 (1)				
Gly	1.00 (1)	1.41 (1)	1.00 (1)	1.15 (1)	0.96 (1)	1.19 (1)
Ala			1.02 (1)			0.13
Cys/2						
Val	0.14	0.14		0.12		
Met			0.67[c] (1)			
Ile	0.96 (1)	1.03 (1)	1.86 (2)	0.12		
Leu		0.18	0.10			
Tyr	0.71 (1)	0.66 (1)				
Phe		0.14	1.24 (1)			
Totals	(7)	(6)	(7)	(5)	(4)	(5)
Electrophoretic migration (cm)[d]	+7.5	+3.7	+3.4	+0.3 (N)	−4.4	+0.3 (N)
Color reactions[e]	Pauly	Pauly	Chloroplatinate	Sakaguchi	Sakaguchi	Sakaguchi
% yield	12	2.1	33	10	51	4.8
Purification procedure[f]		GF, PE	PC		PE	PE

	T-20 (Residues 92–97)	T-21 (Residues 93–97)	T-22 (Residues 98–99)	T-23 (Residues 98–100)	T-24 (Residues 100–104)	T-25 (Residues 101–104)
Residue[b]						
Lys	1.37 (1)		1.15 (1)	1.97 (2)	1.28 (1)	
His						
Arg						
Asx	1.00 (1)	0.95 (1)				
Thr	0.21				0.93 (1)	1.02 (1)
Ser	0.21		0.10	0.24	1.81 (2)	1.80 (2)
Glx					1.04 (1)	1.06 (1)
Pro						
Gly	0.18		0.11	0.20		
Ala	0.88 (1)	1.00 (1)				
Cys/2						
Val	0.17					
Met						
Ile	0.84 (1)	0.98 (1)				
Leu	1.00 (1)	1.05 (1)	0.85 (1)	1.02 (1)		
Tyr	0.75 (1)	0.68 (1)				
Phe						
Totals	(6)	(5)	(2)	(3)	(5)	(4)
Electrophoretic migration (cm)[d]	+0.2 (N)	−4.2	+6.2	+11.6	+0.2 (N)	−5.6
Color reactions[e]	Pauly	Pauly			Yellow	
% yield	2.4	46	32	22	17	55
Purification procedure[f]	GF, PE		PE	PE		

ninhydrin color. Reactions with specific staining reagents were as indicated; Pauly, Sakaguchi, Ehrlich, and chloroplatinate indicate a positive test on a peptide map for Tyr and/or His, Arg, Trp, and sulfur-containing residues, respectively. [f] Purification procedures in addition to the initial fractionation by ion-exchange chromatography (Figure 2), if any, are indicated. Abbreviations are: FFE, free-flow electrophoresis; GF, gel filtration on a Sephadex G-25 column; PC, preparative paper chromatography; and PE, preparative paper electrophoresis.

TABLE III: Amino Acid Compositions of "Insoluble" Tryptic Peptide Fraction (T-12 and T-13) and Derived Peptides.[a]

	T-12 (Residues 56–72) and T-13 (Residues 56–73) Mixture[b]		T-12-Cl (Residues 56–59)[b]	T-12-C2 (Residues 60–65)	T-12-C3 (Residues 66–67)	T-12-C4 (Residues 66–68)
	24 hr	96 hr	24 hr			
Residue						
Lys	1.49	1.16 (1–2)				
Asx	2.00	1.91 (2)		0.97 (1)		
Thr	0.97	0.69 (1)		1.03 (1)		
Ser			0.26		0.30	0.31
Glx	3.20	3.17 (3)		2.00 (2)		
Pro	0.83	0.98 (1)				
Gly	1.14	0.95 (1)	1.00 (1)		0.23	0.27
Val	1.54	1.95 (2)	0.48 (1)		1.00 (1)	1.02 (1)
Ile	0.57	1.03 (1)	0.52 (1)			
Leu	1.89	1.87 (2)		1.01 (1)		0.98 (1)
Tyr	0.86	0.97 (1)			0.73 (1)	0.71 (1)
Phe	0.97	0.98 (1)		0.96 (1)		
Trp		(1)	(1)			
Totals		(17–18)	(4)	(6)	(2)	(3)
Electrophoretic migration (cm)	Not determined		+0.5 (N)	−5.2	+0.7 (N)	+0.4 (N)
Color reactions[c]	Not determined		Yellow, Ehrlich		Pauly	Pauly

	T-12-C5 (Residues 66–72)	T-13-C5 (Residues 66–73)	T-13-C6 (Residues 68–73)	T-13-C7 (Residues 69–73)
Residue				
Lys	0.94 (1)	1.87 (2)	2.11 (2)	1.90 (2)
Asx	1.02 (1)	1.12 (1)	1.06 (1)	1.08 (1)
Thr				0.13
Ser		0.36	0.11	0.37
Glx	1.07 (1)	1.07 (1)	1.06 (1)	0.95 (1)
Pro	0.99 (1)	0.65 (1)	0.87 (1)	1.01 (1)
Gly		0.30	0.15	0.37
Val	1.02 (1)	0.91 (1)		
Ile				
Leu	0.99 (1)	0.97 (1)	0.91 (1)	
Tyr	0.56 (1)	0.75 (1)		
Phe				
Trp				
Totals	(7)	(8)	(6)	(5)
Electrophoretic migration (cm)	+0.8 (N)	+3.8	+4.2	+6.4
Color reactions	Pauly	Pauly		

[a] Footnote *a* of Table II applies. [b] Compositions of the peptide T-12 and T-13 mixture after 24 and 96 hr of acid hydrolysis are given. The results show incomplete release of valine and isoleucine in the 24-hr hydrolysate, which showed about 0.5 residue of each. A 24-hr hydrolysis of peptide T-12-Cl similarly released only about half of the valine and isoleucine. [c] See footnote *e*, Table II.

values for valine and isoleucine in the protein hydrolysates are attributed to the presence of an Ile–Val bond (residues 57 and 58, Figure 3), which hydrolyzes only slowly (see Table III). A tryptophan content of one residue is assigned to the lamprey cytochrome *c* because of the detection of only one tryptophan-containing peptide (T-12-Cl, Table III) with Ehrlich reagent on peptide maps of a tryptic digest of the protein or chymotryptic fragments of tryptic peptides. Peptide T-12-Cl was shown to have a single tryptophanyl residue by sequence analysis.

The compositions of peptides isolated from a tryptic digest of the cytochrome *c* are given in Tables II and III. Peptides which were soluble in 0.2 N pyridine–acetate (pH 3.2), referred to as the "soluble" tryptic peptides (Table II), were isolated by ion-exchange chromatography (Figure 1). Some peptides required further purification as indicated in Table II. A fraction of the tryptic digest insoluble in the pyridine–acetate buffer, designated the "insoluble" tryptic peptides, and peptides isolated from a chymotryptic digest of the insoluble fraction had the compositions given in Table III. The latter peptides were separated by ion-exchange chromatography (Figure 2). Peptide T-12-C3 was further purified by paper chromatography. These compositions and subsequent sequence analysis of the chymotryptic peptides showed that the insoluble fraction contained two peptides (T-12 and T-13) representing residues 56–72 and 56–73, respectively (Figure

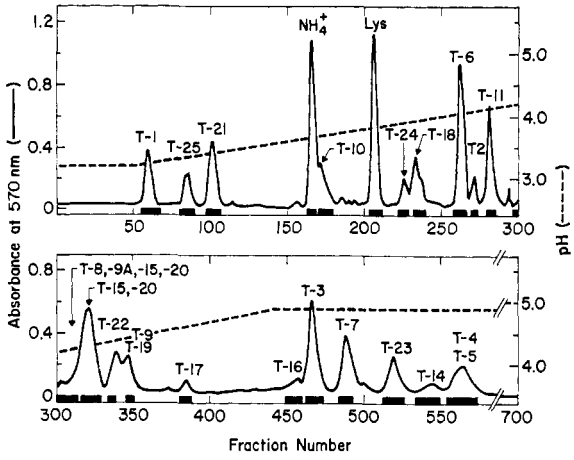

FIGURE 1: Chromatographic separation of the soluble peptides from a tryptic digest of 200 mg of lamprey cytochrome *c* on a 0.9 × 160 cm column of AG 50W-X2. The column was eluted at 40° at the rate of 20–25 ml/hr first with a linear gradient from 0.2 M pyridine–acetate (pH 3.2) to 1.0 M pyridine–acetate (pH 5.0) over a volume of 1000 ml and then with a linear gradient from the latter buffer to 2.0 M pyridine–acetate (pH 5.0) over 500 ml. Fractions of 2.7 ml were collected and peptides were detected by a ninhydrin assay ($A_{570\ nm}$) using 0.10-ml aliquots of every other fraction for the assay. Fractions were pooled as indicated by the bars along the abscissa.

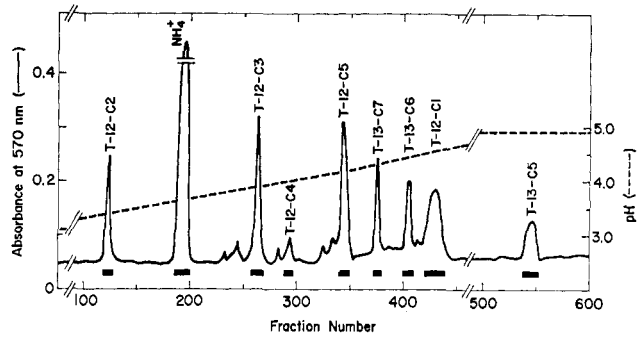

FIGURE 2: Chromatographic separation of peptides from a chymotryptic digest of a mixture of "insoluble" tryptic peptides (T-12 and T-13) from lamprey cytochrome *c* on a 0.9 × 160 cm column of AG 50W-X2. The column was eluted at 40° with a linear gradient from 0.2 M pyridine–acetate (pH 3.1) to 1.0 M pyridine–acetate (pH 5.0) over a volume of 1000 ml with a flow rate of 20–22 ml/hr. Fractions of 2.3 ml were collected, and peptides were detected by ninhydrin assays ($A_{570\ nm}$) on 0.25-ml aliquots of every other fraction. Fractions were pooled as indicated by the bars along the abscissa.

3), and differing only in that T-13 had an additional lysyl residue (residue 73).

The amino acid sequence of the cytochrome (Figure 3) was determined from the sequences of the tryptic peptides, the relative positions of these peptides being deduced by analogy with the structures of cytochromes *c* from other eukaryotic species (see Wojciech and Margoliash, 1970; Dickerson, 1972; and Dayhoff, 1972). The amino acid sequences of the peptides were determined by Edman degradation without further modification as indicated in Figure 3, with the exceptions of peptides T-1, T-5, T-9, T-12, and T-13. These five peptides presented special problems and the data relating to their sequences are summarized below.

Peptide T-1 (Residues 1–5). This peptide, which was acidic, was not degradable by the Edman method, but was ninhydrin-positive, as expected from the presence of a lysyl residue. A ninhydrin-negative component, aspartic acid, glycine, and peptide T-1A (Val, 1.07 [1]; Glu, 1.07 [1]; Lys, 0.87 [1]) were isolated after partial acid hydrolysis of 3.2 μmol of the peptide in 3 ml of 0.1 N acetic acid in a sealed glass tube at 105° for 7 hr. The components were separated by chromatography on an AG 50W-X2 column (0.9 × 60 cm) eluted at 40°, first with 0.2 M pyridine–acetate (pH 3.1) (70 ml), and then with a linear gradient between the initial buffer (250 ml) and 1.0 M pyridine–acetate (pH 5.0) (250 ml); they eluted in the order listed above and were detected with ninhydrin reagent after alkaline hydrolysis. The glycine and aspartic acid were identified by comparison with the authentic compounds by paper electrophoresis at pH 6.5 and paper chromatography in 1-butanol–acetic acid–water (200:30:75) after detection with ninhydrin reagent. The ninhydrin-negative component, which was detected on paper with a hypochlorite–starch–iodide reagent (Pan and Dutcher, 1956), was indistinguishable from authentic *N*-acetylglycine in these same two analytical systems and was assumed to be identical with it on that basis. The sequence of T-1A was established by three steps of Edman degradation. T-1A was electrophoretically neutral at pH 6.5, in accord with the identification of a glutamyl residue in it.

The assigned sequence of peptide T-1 (Figure 3) is consistent with its lack of susceptibility to Edman degradation, the known susceptibility of bonds on either side of aspartyl (or asparaginyl) residues to hydrolysis by weak acids, and the specificity of trypsin. The electrophoretic mobility of the peptide indicated that it contained an aspartyl as well as a glutamyl residue, giving it a net negative charge of two, as expected from its composition and the fact that the α-amino group is blocked. It should be emphasized, however, that the methods of identifying the amino-terminal residue do not completely preclude the possible presence on the glycine of an amino-blocking group other than an acetyl moiety.

Peptide T-5 (Residues 14–22). The sequence of this heme-containing nonapeptide was determined by eight steps of Edman degradation after removal of the heme by reductive cleavage and carboxymethylation of the resulting cysteinyl residues as described in Materials and Methods. Lysine was assigned to the C-terminus from trypsin specificity. His> PhNCS was not identified in the fifth degradation step; histidine was assigned to this position by difference, from the composition of the peptide.

Peptide T-9 (Residues 39–53). Edman degradation established the sequence of the first 13 residues except for the identity of the 11th residue, the PhNCS derivative of which was not detected. Since only a small amount of this peptide was isolated, peptide T-10 was used for further studies. The latter peptide had the same composition as T-9 except that it contained only one rather than two lysyl residues. Three peptides, T-10-C1 (Thr, 0.95 [1]; Glu, 1.05 [1]; Pro, 0.85 [1]; Gly, 2.05 [2]; Ala, 1.05 [1]; Phe, 0.95 [1]; Ser, 0.21), T-10-C2 (Ser, 1.04 [1]; Tyr, 0.96 [1]), and T-10-C3 (Lys, 0.99 [1]; Asp, 2.00 [2]; Thr, 0.95 [1]; Ala, 1.06 [1]; Ser, 0.32) were isolated from a chymotryptic digest of 3 μmol of T-10 by chromatography on a 0.9 × 60 cm column of AG 50W-X2 in pyridine–acetate buffers as described for chromatography of the partial acid hydrolysate of peptide T-1, above. Edman degradation established the sequence of T-10-C3, which, on the basis of its composition and the specificity of trypsin represents the C-terminal segment of peptides T-9 and T-10, thus establishing the complete sequence of T-9. The assigned sequence is consistent with the electrophoretic mobility of the peptide (Table II). As seen in Table II, a small amount of a peptide (T-9A) with the same amino acid composition as T-9

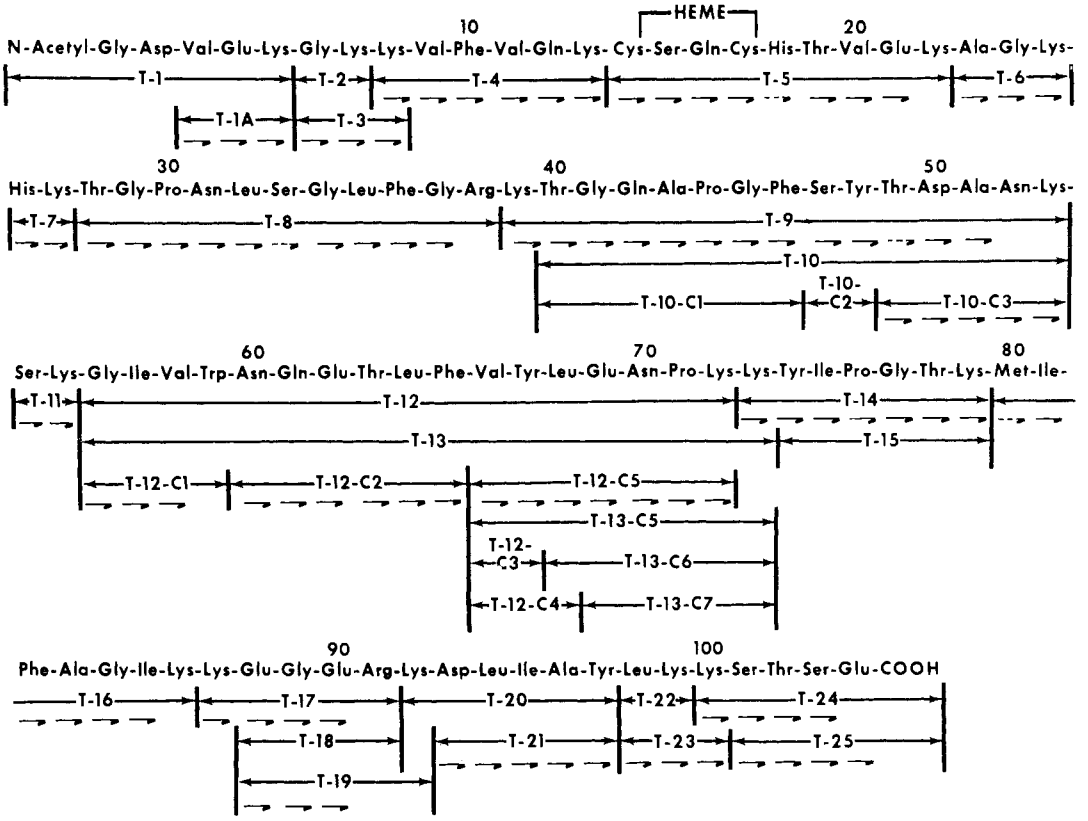

FIGURE 3: The amino acid sequences of the tryptic peptides, indicated by double-headed arrows, and the proposed sequence of the lamprey cytochrome *c*: (→) indicates residues placed by Edman degradation; (⇢) denotes residues for which the PhNCS derivative was not identified. Of these, Thr-49 (in peptide T-9) was placed by sequence analysis of a related peptide, T-10, as shown; His-18, Ser-33, Met-80, *N*-acetyl Gly-1, and Asp-2 were located in the assigned positions as described in the text.

but different electrophoretic mobility was also isolated. The peptide eluted from the cation-exchange column earlier than T-9 (Figure 1) and is presumably a deamidated form of T-9.

Peptides T-12 (Residues 56–72) and T-13 (Residues 56–73). The compositions of peptides isolated from a chymotryptic digest of a mixture of the "insoluble" peptides T-12 and T-13 by ion-exchange chromatography (Figure 2) are given in Table III with the composition of the mixture of T-12 and T-13. The sequences of the chymotryptic peptides were determined by Edman degradation and their order in the parent peptides (Figure 3) was assigned by analogy with the sequences of cytochromes *c* from other eukaryotic species. Peptide T-12-C5 was put at the C-terminus of T-12 since both peptides have a single lysyl residue, while T-13-C5 was assigned to the C-terminus of T-13, both of which contain two lysines. All the other chymotryptic peptides were derived from both T-12 and T-13 but are arbitrarily designated as products of T-12 (Table III, Figures 2 and 3). The sequence of T-13-C5 was not determined; it was assumed by analogy with T-12-C5 and from the specificity of trypsin. The tryptophanyl residue of T-12-C1 was placed at the C-terminus from the specificity of chymotrypsin and by difference from the amino acid compositions since all other residues were accounted for.

The PhNCS derivatives of Ser-33 (peptide T-8) and Met-80 (peptide T-16) were not identified. These residues were given the positions indicated from the compositions of the peptides since all other residues were located by Edman degradation, except for the C-terminal residues (arginine and lysine, respectively), which were placed on the basis of trypsin specificity.

Residues 8–13 of the protein are represented in Figure 3 and Table II by only a single peptide, T-4, of low yield (18%). Although a pentapeptide representing residues 9–13, predicted from the sequence of the cytochrome on the basis of trypsin specificity, was not isolated in purified form, amino acid analysis clearly indicated that this peptide eluted in the fraction containing T-15 and T-20 in the initial ion-exchange chromatography step (Figure 1). Peptide T-20 was separated from the other two peptides by gel filtration on Sephadex G-25, but T-15 (a hexapeptide) and the pentapeptide coeluted in yields of about 44 and 28%, respectively, and were not adequately resolved by rechromatography on a column of AG 50W-X8 or by peptide mapping. Since T-4 spans the entire segment of the protein represented by the pentapeptide and since a small amount of T-15 was obtained free of the pentapeptide in the initial fractionation step (Figure 1 and Table II) and T-14 spans the entire segment represented by T-15, no further attempt was made to separate the mixture.

Discussion

Even though the cyclostomes are descendants of a very primitive group of vertebrates which antedate the development of a lower jaw, the cytochrome *c* of a present day cyclostome, the Pacific lamprey, *Entosphenus tridentatus*, is clearly a typical vertebrate cytochrome. As in other vertebrate cytochromes *c*, the chain of 104 residues starts with the characteristic *N*-acetylglycine. This is the position in which, for all nonvertebrate cytochromes *c* examined to date, the acetyl is replaced by a short chain of amino acids up to nine residues long, and in the proteins of higher plants the acetyl recurs at

the N-terminus of the extra segment (see Margoliash, 1972; Dayhoff, 1972). The typical unvaried metazoan sequence of residues 70–80 is conserved, as are the residues which have been assigned functional roles on the basis of the spatial structure of the protein in the ferric and ferrous forms (Dickerson *et al.*, 1971; Takano *et al.*, 1973; Margoliash *et al.*, 1972). Among the latter are histidine-18 and methionine-80 which provide the principal side chains bound to the heme iron, tyrosyl residues 67 and 74, which may be involved in the reduction to the ferrous form, tryptophan-59 and tyrosine-48, which make the hydrogen bonds holding the posterior propionyl side chain of the porphyrin in the hydrophobic heme crevice, phenylalanine-82, which swings out and into the protein on oxidation–reduction, and the two invariant arginines, residues 38 and 91, of so far unknown function. The lamprey protein sequence has the clusters of hydrophobic and basic residues typical of other eukaryote cytochromes *c* (Margoliash and Smith, 1965; Margoliash and Fitch, 1968). It functionally repletes cytochrome *c* depleted rat liver mitochondria in a manner which is quantitatively indistinguishable from that of the numerous other cytochromes *c* tested (V. Byers, C. H. Kang, and E. Margoliash, unpublished experiments).

Excluding the rattlesnake protein (Bahl and Smith, 1965), the amino acid sequence of which appears sufficiently peculiar to place it in an unexpected position in a statistical phylogenetic tree (Fitch and Margoliash, 1967, 1970), lamprey cytochrome *c* differs from the cytochromes *c* of other vertebrates by 14–21 residues, of insects by 24–28 residues, of higher plants by 40–43 residues, of fungi by 42–47 residues, and of protozoans by 43–48 residues. This degree of variability is well within the range of differences among other vertebrate cytochromes *c* and indeed a statistical phylogenetic tree, based on the amino acid sequences of this homologous set of proteins and developed according to the procedure of Fitch and Margoliash (1967), places the lamprey within the expected taxonomic group of species (see Fitch and Margoliash, 1970). From that phylogeny, somewhat unexpected relationships of the cyclostomes to the higher vertebrates might be inferred. The cyclostome (Pacific lamprey), elasmobranch or cartiloginous fish (dogfish) and teleosts or bony fish (tuna, bonito, and carp), are apparently more closely related to each other than to the higher vertebrates. This would seem to imply that the higher vertebrates diverged from the three groups of lower vertebrates, before the latter diverged from each other, in contrast with the common view that the cyclostomes gave rise to the cartilaginous fish, which gave rise to the bony fish, from which ultimately the higher vertebrates arose. However, experience suggests that conclusions based on large phyletic distances, together with representation from only a few taxa (one cyclostome and one elasmobranch), must be considered tentative until supported by further appropriate amino acid sequences. Nevertheless, it must be emphasized that when such data are obtained a critical test will be available of any hypothesis concerning the phylogenetic relations of the lower vertebrates and the higher vertebrate classes, and that this tool can be refined nearly indefinitely by the examination of other sets of orthologous proteins.

One of the difficulties inherent in the accumulation of the large number of amino acid sequences for an orthologous protein from different species is the tendency for workers to become less rigorous in their amino acid sequence determinations, as the structures of more and more variants are worked out. A common shortcut is to assume that peptides having the same amino acid composition have the same amino acid sequence. The danger of such a procedure has previously been emphasized (Margoliash and Fitch, 1969), giving as example the carboxyl-terminal sequences of the cytochromes *c* of two birds, the pigeon and duck, and of the rattlesnake, which are Thr-Ala-Ala-Lys, Ala-Thr-Ala-Lys, and Lys-Thr-Ala-Ala, respectively. A similar case occurs in lamprey cytochrome *c* as the presence of a lysine at position 92 makes it possible to obtain two tryptic peptides from the sequence Lys-Lys-Glu-Gly-Glu-Arg-Lys (residues 86–92), having the same composition but with the lysyl residue at opposite ends. These are peptide T-17 comprising residues 87–91, and peptide T-19 containing residues 88–92. It is interesting to note that these two peptides separated well by ion-exchange chromatography on a sulfonated polystyrene resin (see Figure 1). The tendency to assign to both peptides the sequence of T-17 is greatly strengthened by the fact that a lysine in position 92 occurs among the vertebrate proteins only in lamprey and bullfrog cytochromes *c*, while the lysine in position 87 is very common. Thus, identical compositions even for peptides coded for by closely related orthologous genes (Fitch and Margoliash, 1970) or by the same gene, as in the present case, are no guarantee of identity of sequence. Moreover, when the amino acid sequences of peptides are decided by analogy and the conclusions are used to infer phylogeny the reasoning becomes grossly circular since in the first place the amino acids in the peptides were aligned to maximize assumed homologous relationships.

An interesting point of hydrolysis by the Tos-PheCH$_2$Cl treated preparation of trypsin employed was at the Tyr–Leu bond between residues 97 and 98. Peptides T-20 and T-21 resulting from that split were obtained in a good overall yield of 48% and no peptide overlapping the Tyr-Leu sequence was obtained. No other chymotryptic-like hydrolyses were observed. A similar situation was previously reported for kangaroo cytochrome *c* (Nolan and Margoliash, 1966) and hydrolyses at the same bond were observed in tryptic digests of moth (Chan and Margoliash, 1966a), chicken (Chan and Margoliash, 1966b), and human (Matsubara and Smith, 1963) cytochromes *c*. These have been attributed to contamination of the trypsin with chymotrypsin but this explanation becomes less tenable in the present case as no other hydrolyses attributable to chymotrypsin were detected and the trypsin preparation had been treated with a chymotrypsin-specific inhibitor. It would seem that ordinary trypsin specificity may have been affected by the spatial structure of the protein in the region of this bond and by side chains other than those immediately adjacent to the bond split.

References

Ambler, R. P. (1963), *Biochem. J. 89*, 349.

Bahl, O. P., and Smith, E. L. (1965), *J. Biol. Chem. 240*, 3585.

Chan, S. K., and Margoliash, E. (1966a), *J. Biol. Chem. 241*, 335.

Chan, S. K., and Margoliash, E. (1966b), *J. Biol. Chem. 241*, 507.

Crestfield, A. M., Moore, S., and Stein, W. H. (1963), *J. Biol. Chem. 238*, 622.

Dayhoff, M. O. (1972), Atlas of Protein Sequence and Structure, Vol. 5, Silver Spring, Md., National Biomedical Research Foundation.

Dickerson, R. E. (1971), *J. Mol. Evol. 1*, 26.

Dickerson, R. E. (1972), *Sci. Amer. 226*, 58.

Dickerson, R. E., Takano, T., Eisenberg, D., Kallai, O. B., Samson, L., Cooper, A., and Margoliash, E. (1971), *J.*

Biol. Chem. 246, 1511.
Fitch, W. M., and Margoliash, E. (1967), *Science 155*, 279.
Fitch, W. M., and Margoliash, E. (1970), *in* Evolutionary Biology, Vol. 4, Dobzhansky, Th., Hecht, M. K., and Steere, E. C., Ed., New York, N. Y., Meredith Corp., p 67.
Hirs, C. H. W., Moore, S., and Stein, W. H. (1956), *J. Biol. Chem. 219*, 623.
Jeppson, J., and Sjöquist, J. (1967), *Anal. Biochem. 18*, 264.
Kostka, V., and Carpenter, F. H. (1964), *J. Biol. Chem. 239*, 1799.
Margoliash, E. (1972), *Harvey Lect. 66*, 177.
Margoliash, E., Dickerson, R. E., and Adler, A. D. (1972), *in* The Molecular Basis of Electron Transport, Schultz, J., and Cameron, B. F., Ed., New York, N. Y., Academic Press, p 153.
Margoliash, E., and Fitch, W. M. (1968), *Ann. N. Y. Acad. Sci. 151*, 359.
Margoliash, E., and Fitch, W. M. (1969), *Nat. Acad. Sci. Nat. Res. Counc., Publ. No. 1692*, 357.
Margoliash, E., Fitch, W. M., and Dickerson, R. E. (1968), *Brookhaven Symp. Biol. 21*, 259.
Margoliash, E., and Smith, E. L. (1965), *in* Evolving Genes and Proteins, Bryson, V., and Vogel, H. J., Ed., New York, N. Y., Academic Press, p 221.
Margoliash, E., and Walasek, O. F. (1967), *Methods Enzymol. 10*, 339.
Matsubara, H., and Smith, E. L. (1963), *J. Biol. Chem. 238*, 2732.
Nolan, C., and Margoliash, E. (1966), *J. Biol. Chem. 241*, 1049.
Nolan, C., and Margoliash, E. (1968), *Annu. Rev. Biochem. 37*, 727.
Nolan, C., Margoliash, E., Peterson, J. D., and Steiner, D. F. (1971), *J. Biol. Chem. 246*, 2780.
Pan, S. C., and Dutcher, J. D. (1956), *Anal. Chem. 28*, 836.
Roseau, G., and Pantel, P. (1969), *J. Chromatogr. 44*, 392.
Takano, T., Kallai, O. B., Swanson, R., and Dickerson, R. E. (1973) (in press).
Wojciech, R., and Margoliash, E. (1970), *in* Handbook of Biochemistry, Selected Data for Molecular Biology, 2nd ed, Sober, H. A., Ed., Cleveland, Ohio, The Chemical Rubber Co., pp C228–C233.

Some Properties of Cross-Linked Polymers of Glutamic Dehydrogenase†

Robert Josephs, Henryk Eisenberg, and Emil Reisler*

ABSTRACT: Polymers of glutamic dehydrogenase are shown to be stabilized by cross-linking with glutaraldehyde. The nature of the cross-linked material is dependent upon the conditions chosen for the cross-linking reaction and appropriate conditions yield cross-linked polymers which are enzymatically active. Gel filtration of such polymers results in separation of the cross-linked material into fractions of narrow size distribution with molecular weights as high as 3×10^6. The molecular weights of such fractions are invariant with concentration in contrast with the behavior of the native enzyme. Acrylamide gel electrophoresis in sodium dodecyl sulfate shows that the major fraction of the cross-links bridge regions within a single molecule rather than between molecules. Kinetic parameters for the enzymatic reaction determined for fractions of different molecular weights show that enzymatic activity of the fixed enzyme is independent of the degree of polymerization.

Glutamic dehydrogenase is known to undergo polymerization (Sund, 1968) to form linear polymers (Eisenberg and Tomkins, 1968). The reacting species are in rapid equilibrium and their size distribution depends upon the concentration of enzyme present. At low protein concentration (less than 0.05 mg/ml) the monomer (mol wt 320,000) is by far the predominant species and, as the concentration is raised, progressively higher *n*-mers are generated; under appropriate conditions linear polymers up to 15–20 molecules long can be formed (for summaries, see Josephs *et al.*, 1972; Eisenberg, 1971). The distribution of sizes is not narrow but rather a broad spectrum of particle lengths is obtained (Reisler *et al.*, 1970). As an abundant and controversial literature will attest, these characteristics of the polymerization reaction introduce formidable interpretative difficulties in studies of the physical state of the enzyme.

An additional element of complexity is introduced upon examination of the relationship between the polymerization reaction and enzymic activity. This is exemplified by the observation that the presence of 10^{-3} M NADH and certain steroid agents or 10^{-3} M GTP cause nearly complete depolymerization and parallel inhibition of enzymatic activity. These and similar observations, when taken in conjunction, have led various workers to consider the relationship between enzymatic activity and polymerization and the role the depolymerizing agents mentioned above may play in regulating the form and activity of glutamic dehydrogenase *in vivo* (see, for example, reviews by Frieden, 1963a,b; Stadtman, 1966; and Tomkins *et al.*, 1963, 1965). A direct attack on this question is difficult since enzymatic activity is generally measured at concentrations so low that essentially all the enzyme is dissociated into monomer.

† From the Department of Polymer Research, The Weizmann Institute of Science, Rehovot, Israel. *Received May 15, 1973.*