THE AMINO ACID SEQUENCE OF A CYTOCHROME *c* FROM A PROTOZOAN *CRITHIDIA ONCOPELTI*

G.W. PETTIGREW

Department of Molecular Biology and Department of Biochemistry, University of Edinburgh, Edinburgh EH9 3JR, Scotland

Received 8 February 1972

1. Introduction

Eukaryotic cytochromes *c* form a homologous family of proteins with considerable conservation of primary structure. The haem binding site follows the pattern

–Basic–Cys–X–Y–Cys–His

in all of the cytochromes *c* studied from higher organisms. Some work has been done on the unicellular eukaryotes, protozoa and algae [1–3], but no sequence has been published. Two kinds of *c*-type cytochromes are distinguished, those presumed to be members of the above group of mitochondrial proteins and those presumed to be connected with photosynthesis e.g. cytochrome *C*-552 of Euglena. Both groups may be interesting for the attempts to construct molecular phylogenies and also in the relation of structure with function and properties.

2. Experimental

Crithidia oncopelti is a trypanosomatid – a parasitic flagellate – closely related to the organism studied by Hill [1], *Crithidia fasciculata.* The cytochrome *c* was purified by column chromatography on CM-cellulose (0.04 M phosphate, pH 7.2) running the protein first in the reduced form and then in the oxidised form, a procedure which yields pure protein as judged by chemical, electrophoretic and chromatographic criteria. The protein has distinctive properties and the prosthetic group appears to be different from haem *C* [4].

The sequence work was undertaken in an attempt to discover whether an explanation of these properties might be possible in terms of the primary structure of the protein. The haem was removed by the method of Ambler [5] and tryptic and chymotryptic digestions were performed on 1.5–2 μmole quantities of the apoprotein. Peptides were purified by gel filtration, paper electrophoresis and chromatography according to the methods of Ambler [5].

3. Results and discussion

Fig.1 shows the proposed sequence and summarises the evidence for it. Table 1 shows the amino acid composition as obtained from amino acid analyses and from sequence study.

The following points require emphasis.

The bond between tyrosine 48 and serine 49 was hydrolysed in high yield by trypsin and the evidence for this bond is from considerations of homology only. No N-terminal amino acid could be identified in the whole protein while carboxypeptidase B liberated equimolar lysine and leucine. The N-terminal peptides are, as yet, incompletely characterised. The first two residues are tentatively assigned on the evidence of the tryptic peptide (Ala_1, Pro_1, trimethyllysine$_1$, Arg_1). The N-terminus of this peptide is blocked and thermolysin digestion yields two peptides with compositions (Pro, trimethyllysine) and (Ala, Arg). The presence, in low yield, of the tryptic peptide (Ala-7 to Lys 5) may have been due to incomplete methylation at Lys-8 (trypsin does not hydrolyse trimethyllysine bonds).

Tryptophan was not quantitated in the sequence

North-Holland Publishing Company – Amsterdam

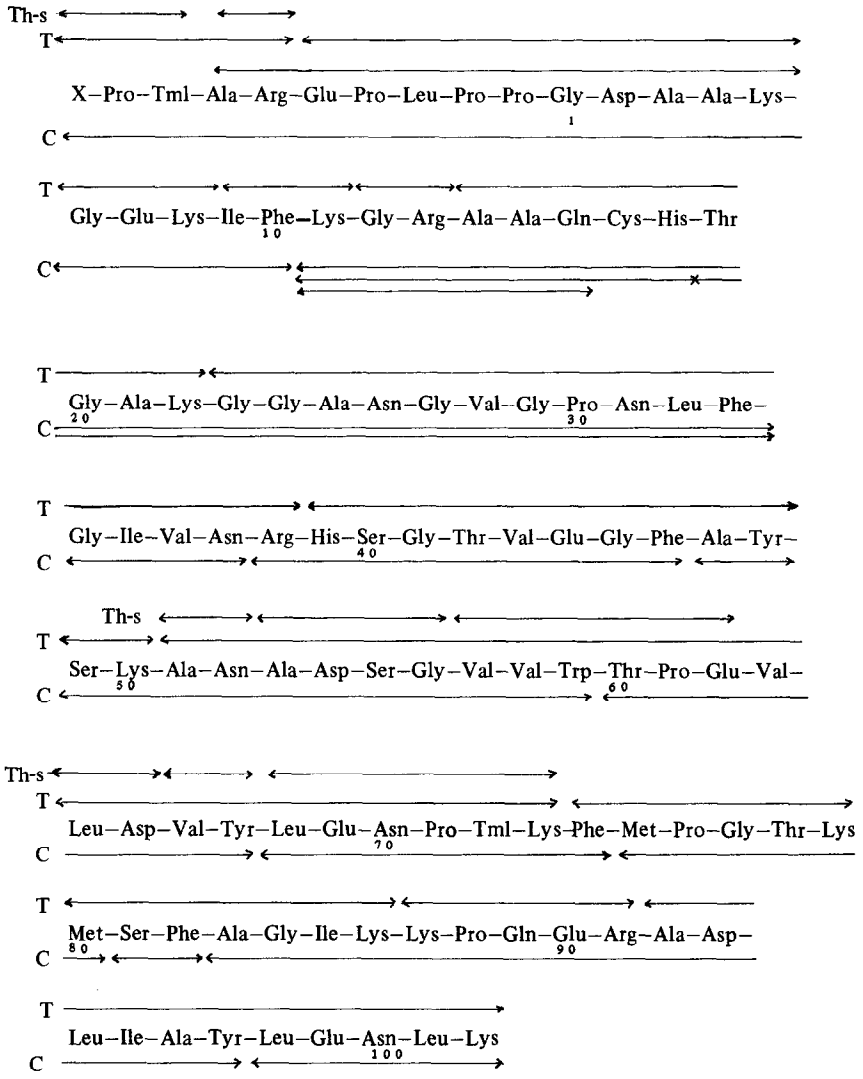


Fig.1. Proposed structure of cytochrome *c* from *Crithidia oncopelti* showing peptides isolated from tryptic (T) and chymotryptic (C) digestions and thermolysin sub-digestions (Th-s). The glycine corresponding to Gly 1 of vertebrate cytochromes *c* is numbered 1. Peptides were sequenced by the dansyl-Edman procedure [10], employing enzymic sub-digestion when necessary. Amides were located by examining the mobilities either of parent peptides or of remainder peptides during Edman degration. Tml = trimethyllysine.

study. Only one Ehrlich positive peptide was found in both chymotryptic and tryptic digests and the large Ehrlich positive tryptic peptide yielded only one Ehrlich positive peptide on sub-digestion with chymotrypsin or thermolysin. The method of Goodwin and Morton [6] gave 3.5 moles tyrosine and 0.9 mole tryptophan per mole.

The proposed structure differs in some respects from other eukaryotic cytochromes *c*. Residue 14 is alanine in place of cysteine, a change which is surprising in a position considered to be invariant and structurally crucial. The N-terminus is extended like that of plant and invertebrate cytochromes and appears to be blocked. Four of the nine prolines occur here as does one of the two trimethyllysine residues. Apart from the alanine substitution in the haem binding

Table 1
Amino acid composition of cytochrome *c (Crithidia oncopelti)* based on residues per atom of iron compared with total residues in the proposed sequence.

	Composition	Sequence
Asp	9.9	10
Thr	4.0	4
Ser	4.0	4
Glu	8.7	9
Pro	9.1	9
Gly	13.6	14
Ala	12.4	13
Val	6.7	7
Met	1.8	2
Ile	3.9	4
Leu	6.7	7
Tyr	2.7	3
Phe	5.0	5
His	1.9	2
Lys	9.8	10
Tml	1.8	2
Arg	3.9	4
Cys	1.1	1
Trp	–	1
	Total	111

Aliquots of the same solution were used for 20 hr and 70 hr hydrolyses and for iron determination. Cysteic acid was determined separately after removal of the heam and performic acid oxidation. Tml = trimethyllysine.

site three other "invariant" residues are different: Lys 27 is replaced by a Gly; Tyr 74 by a Phe; and Ile 75 by a Met. Thus out of the 35 "invariant" residues of the cytochrome *c* family [7] 31 are identical in the Crithidia protein and occur in the same linear pattern with no insertions or deletions (although the existence of a bond between residues 48 and 49 is not proven). The heam is probably bound through the single cysteine, leaving one vinyl side chain free. Thus the protein is homologous with mitochondrial cytochromes *c* but the unusual attachment of the prosthetic group results in a distinctive spectrum.

Interesting questions that arise are:

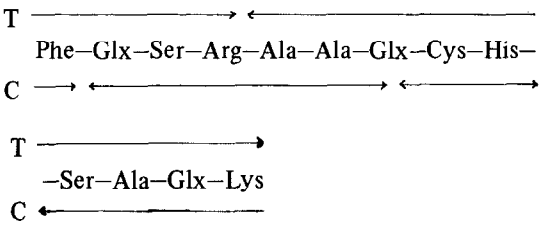
1) Do all protozoan cytochromes *c* have this type of haem attachment?

2) If one of the cysteines is variable, how many truly invariant residues will there be among all cytochromes *c*? Previous estimates may be much too high [1].

3) Are the differences in primary structure reflected in differences in enzymic properties? The properties of other eukaryotic cytochromes *c* studied appear to be remarkably constant insofar as they can be measured [8,9].

Euglena is a unicellular flagellate like Crithidia but is photosynthetic. Its phylogeny is disputed. Two soluble *c*-type cytochromes can be isolated, one of which, Euglena *C*-558, has been purified as above in an attempt to begin to answer questions 1 and 2. The spectral properties are very similar to those of the Crithidial cytochrome. The pyridine haemochrome αpeak is at 553 nm and the test for free vinyl groups was positive.

Preliminary sequence results suggest that the haem binding site is



Acknowledgements

I thank Dr. Richard Ambler for the chance of working in his laboratory and for his help and interest and Drs. Andrew Ryle, John Leaver and Terry Meyer for valuable advice and encouragement.

References

[1] G.C. Hill, Biochim. Biophys. Acta 253 (1971) 78.
[2] T. Yamanaka, Y. Nagata, K. Okunuki, Biochem. J. 63 (1968) 753.
[3] E. Yakushiji, in: Methods in Enzymology, Vol. XXIII, ed. A. San Pietro (Academic Press, New York, 1971) p. 364.
[4] G. Pettigrew and T. Meyer, Biochem. J. 125 (1971) 46P.
[5] R.P. Ambler, Biochem. J. 89 (1963) 349.
[6] T.W. Goodwin and R.A. Morton, Biochem. J. 40 (1946) 628.
[7] W.M. Fitch and E. Margoliash, Biochem. Genet. 1 (1967) 65.
[8] V.S. Byers, D. Lambeth, H.A. Lardy and E. Margoliash, Federation Proc. 30 (1971) 1286.
[9] A. Schejter and R. Margalit, FEBS Letters 6 (1970) 278.
[10] W.R. Gray and J.F. Smith, Anal. Biochem. 33 (1970) 36.