Biochem. J. (1979) **177**, 819–823
Printed in Great Britain

The Amino Acid Sequence of Cytochrome *c'* from the Purple Sulphur Bacterium *Chromatium vinosum*

By R. P. AMBLER,* MARGARET DANIEL,* T. E. MEYER,† R. G. BARTSCH† and M. D. KAMEN†‡
* *Department of Molecular Biology, University of Edinburgh, Edinburgh EH9 3JR, Scotland, U.K., and*
† *Department of Chemistry, University of California at San Diego, La Jolla, CA 92037, U.S.A.*

(*Received* 8 *August* 1978)

An amino acid sequence is proposed for the cytochrome *c'* from the photosynthetic purple sulphur bacterium *Chromatium vinosum* strain D. It is a single polypeptide chain of 131 residues, with haem-attachment cysteine residues at positions 121 and 124. The results discredit an earlier report [Dus, Bartsch & Kamen (1962) *J. Biol. Chem.* **237**, 3083–3093] of a di-haem peptide sequence from this protein. The sequence belongs to the same class as the published *Alcaligenes* and *Rhodospirillum rubrum* cytochrome *c'* sequences, but the resemblance is not close. Detailed evidence for the amino acid sequence of the protein has been deposited as Supplementary Publication SUP 50093 (15 pp.) at the British Library Lending Division, Boston Spa, Wetherby, West Yorkshire LS23 7BQ, U.K., from whom copies may be obtained on the terms given in *Biochem. J.* (1978) **169**, 5.

Cytochromes *c'* constitute a class of electron-transport haem proteins that are found primarily in photosynthetic bacteria. The haem is covalently bound to cysteine residues, so the proteins belong by definition to the cytochromes *c*, although there is no other similarity to the well-characterized cytochrome *c* of mitochondria. Cytochromes *c'* are classed in Type II in a sequence-based classification for haem-*c*-containing proteins (Ambler, 1979).

Chromatium vinosum was one of the earliest known sources of cytochrome *c'* (Bartsch & Kamen, 1960), and much of the earlier characterization of this class of proteins was done with material from this organism (Bartsch *et al.*, 1961; Dus *et al.*, 1962; Taniguchi & Kamen, 1963; Dus & Kamen, 1963). Most cytochromes *c'* have a mol.wt. of 28000 (Cusanovich, 1971), and are composed of identical subunits that each contain one haem group and that are 14000 daltons in size (Cusanovich *et al.*, 1970; Kennel *et al.*, 1972). *Chromatium* cytochrome *c'* is one of the most difficult to dissociate into a monomeric form, and this property may have contributed to early publication of an erroneous sequence of a 'di-haem peptide' (Dus *et al.*, 1962), a finding that has since been discredited (Kennel *et al.*, 1972).

The complete amino acid sequences of two cytochromes *c'* have been reported. They are from a denitrifying *Alcaligenes* species (Ambler, 1973) and from the purple non-sulphur photosynthetic bacterium *Rhodospirillum rubrum* (Meyer *et al.*, 1975). The proteins are similar in length, and show 28% identity if aligned with one four-residue gap and two single-residue gaps.

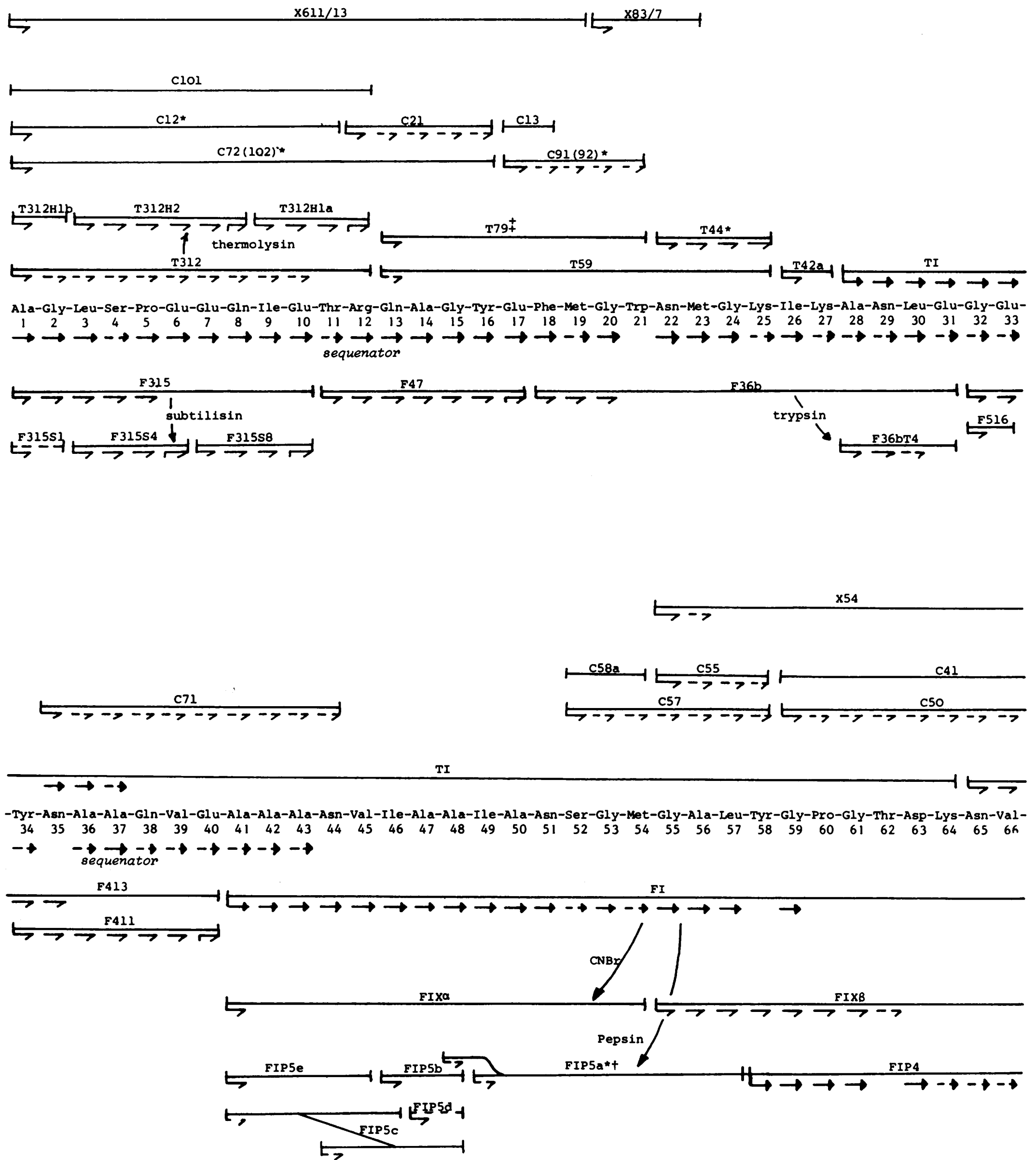
We have determined the amino acid sequence of the *Chromatium vinosum* cytochrome *c'* as part of a programme to correlate structures of bacterial cytochromes with function and phylogeny. We report the sequence of the *Chromatium vinosum* protein now in the hope that it will end controversy about its subunit structure, haem content and amino acid sequence around the haem-binding site. The sequence is from a representative of a third bacterial family, and will help to indicate the limits of variability of this class of protein.

Experimental

Preparation of cytochrome c'

Chromatium vinosum strain D (A.T.C.C. 17899) was grown as described by Bartsch & Kamen (1960). The cytochrome *c'* was isolated by the method of Kennel *et al.* (1972). The purity of the protein was assessed by the ratio of A_{280} to that of the Soret maximum (Fe^{3+} form) at 400nm (less than 0.32) and by homogeneity on gel filtration through Sephadex G-75 and chromatography on DEAE-cellulose. The final preparations were homogeneous by *N*- and *C*-terminal group analysis, and the amino acid composition (Table 1) agreed adequately with the deduced sequence (Fig. 1).

‡ Present address: Chemical Biology Development Laboratory, University of Southern California, Los Angeles, CA 90007, U.S.A.


X611/13
X83/7
C101
C12*
C21
C13
C72(102)*
C91(92)*
T312H1b
T312H2
T312H1a
thermolysin
T79‡
T44*
T312
T59
T42a
TI
Ala-Gly-Leu-Ser-Pro-Glu-Glu-Gln-Ile-Glu-Thr-Arg-Gln-Ala-Gly-Tyr-Glu-Phe-Met-Gly-Trp-Asn-Met-Gly-Lys-Ile-Lys-Ala-Asn-Leu-Glu-Gly-Glu-
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33
sequenator
F315
F47
F36b
subtilisin
trypsin
F516
F315S1
F315S4
F315S8
F36bT4
X54
C58a
C55
C41
C71
C57
C50
TI
-Tyr-Asn-Ala-Ala-Gln-Val-Glu-Ala-Ala-Ala-Asn-Val-Ile-Ala-Ala-Ile-Ala-Asn-Ser-Gly-Met-Gly-Ala-Leu-Tyr-Gly-Pro-Gly-Thr-Asp-Lys-Asn-Val-
34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66
sequenator
F413
FI
F411
CNBr
FIXα
FIXβ
Pepsin
FIP5e
FIP5b
FIP5a*†
FIP4
FIP5d
FIP5c

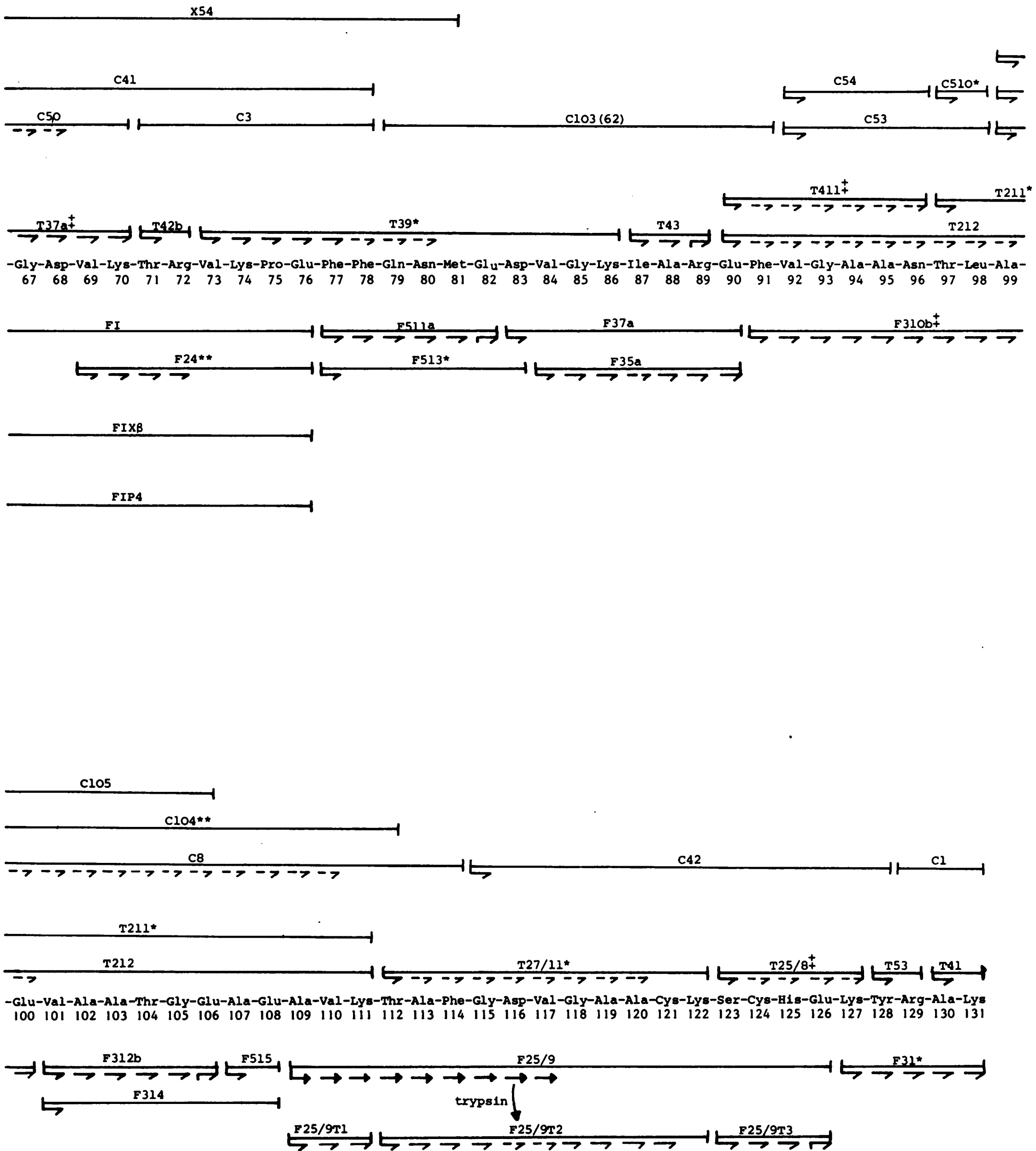


Fig. 1. *Amino acid sequence of Chromatium vinosum strain D cytochrome c′*

Peptides derived by digestion with trypsin (T) or chymotrypsin (C) or by cleavage with CNBr (X) are shown above the sequence, and by digestion with *Staphylococcus aureus* proteinase (F) below the sequence. Peptides from subdigests are labelled with a second capital letter indicating the secondary method of cleavage (H, thermolysin; S, subtilisin; P, pepsin). All peptides were analysed quantitatively for amino acid composition, but those marked * gave substandard and ** particularly bad analyses. Under peptide lines, ⇁ indicates end groups and subsequent residues revealed by phenyl isothiocyanate degradation and identified by the dansyl method, substandard if shown - ⇁. The results for the sequences of chymotryptic peptides are only shown as substandard because the absence of subsidiary dansyl amino acids were not specifically noted in the records. *C*-Terminal residues of peptides identified as free amino acids after removal of the remainder by phenyl isothiocyanate degradation are shown ↱. Peptides marked † were examined by carboxypeptidase A digestion and those marked ‡ by aminopeptidase M digestion. Under the sequence, → indicates residues identified as phenylthiohydantoins after degradation of the whole protein in a sequenator, only tentatively identified if marked - →. The same symbols under peptide lines indicate the results of sequenator degradations on peptides.

Amino acid sequence determination

The amino acid sequence was determined by the general methods used for other bacterial cytochromes *c* (Ambler & Wynn, 1973; Ambler, 1973), and to similar standards. Protein (2–4 μmol) was treated with $HgCl_2$ in 8 M-urea/0.1 M-HCl at 37°C for 16h to remove the haem moiety, and after gel filtration and freeze-drying the sample was digested with a proteinase. The peptides were fractionated by gel filtration followed by high-voltage paper electrophoresis and chromatography, and then analysed quantitatively for amino acid composition and purity. Peptide sequences were investigated by the dansyl/phenyl isothiocyanate method, and by exo- and endopeptidase digestion. The *N*-terminal sequences of the whole protein and of some large peptides were investigated with an automatic sequenator (Beckman model 890A; Edman & Begg, 1967) using a variety of programs. Amide groups were assigned from peptide electrophoretic mobilities, exopeptidase analysis and direct identification of phenylthiohydantoins.

Table 1. *Amino acid composition of Chromatium vinosum cytochrome c′*

Results are shown as residues per molecule. Previous results are from: (1) Kennel *et al.* (1972); (2) Bartsch *et al.* (1961). The latter are recalculated assuming the protein to contain one haem moiety. Results from the present investigation are for protein hydrolysed at 105°C under vacuum. Samples (3) and (4) were of protein treated to remove the haem, and were hydrolysed with 6 M-HCl for 24h and 96h respectively. Sample (5) was of native protein, hydrolysed with 3 M-mercaptoethanesulphonic acid for 96h.

	Previous results		The present investigation			
	(1)	(2)	(3)	(4)	(5)	Sequence
Glycine	16	19	14.7	14.3	13.2	15
Alanine	26	29½	23.1	22.8	22.0	24
Valine	10	12½	8.8	9.3	8.2	10
Leucine	4	5	4.4	4.4	4.2	4
Isoleucine	5	6½	4.3	4.9	4.0	5
Serine	3	2	2.8	2.5	3.0	3
Threonine	6	6	5.9	5.2	5.5	6
Aspartic acid	13	15	12.0	11.7	11.7	4
Asparagine						8
Glutamic acid	20	22½	17.9	17.3	16.8	13
Glutamine						5
Phenylalanine	5	3½	4.8	4.9	4.5	5
Tyrosine	4	6	3.8	3.7	3.5	4
Tryptophan					1.1	1
Cysteine	(2)	(2)	1.4			2
Methionine	4	5	4.0	3.6	3.1	4
Proline	3	4	3.1	3.0	3.6	3
Lysine	10	12	9.1	9.8	9.9	10
Histidine	1	1	1.2	1.2	1.1	1
Arginine	4	4½	4.0	3.8	4.0	4
Total						131

Results

The amino acid composition of the protein is shown in Table 1. The evidence for the amino acid sequence is summarized in Fig. 1. Details of the purification, analysis and sequence-determination experiments on all the peptides shown in Fig 1, and of the automatic-sequenator runs are given in the Supplementary Publication (SUP 50093). The criteria for satisfactory results, and the nature and format of the Supplementary Publication, are given in previous papers (Ambler & Wynn, 1973; Ambler, 1973, 1975).

The amino acid sequence was determined by the characterization of peptides from two tryptic, two staphylococcal-proteinase and one chymotryptic experiment, and from partial characterization of the simpler fragments from a CNBr experiment. The second experiment with staphylococcal proteinase was only to obtain a further quantity of the large peptide FI (Fig. 1).

Discussion

Accuracy of proposed sequence

All the peptides are linked into a single sequence by satisfactory overlaps. The weakest overlap is at residues 89, 90 and 91, where the chymotryptic peptides C103 and C62 were obtained in low yield, probably because of pyrrolidonization of the glutamine-79 residue.

The region of sequence from residue 41 to residue 62 contains no amino acid with a side chain that is ionized at normal pH values. This region occurred in large 'core' peptides (TI and FI) in tryptic and staphylococcal-proteinase digests, and these large peptides could not be subdigested to give convenient peptides in good yield for sequence determination. The sequence for this region shown in Fig. 1 depends on automatic-sequenator results supported by good quantitative amino acid analyses of the large peptides. We consider that the evidence establishes the proposed sequence, but nevertheless recommend that if this protein should be studied by an independent method (such as by X-ray crystallography), especial attention should be paid to this region.

Comparison with other cytochrome c′ sequences

The amino acid sequence of *Chromatium vinosum* cytochrome *c′* is appreciably more similar to the sequence of the protein from the denitrifying *Alcaligenes* species (Ambler, 1973) than it is to the protein from the photosynthetic *R. rubrum* (Meyer *et al.*, 1975).

The 'di-haem peptide'

The present results establish that *Chromatium vinosum* cytochrome *c′* is made up of two identical subunits of about 14000 mol.wt., each containing a single haem group. This finding was anticipated by Kennel *et al.* (1972) on the basis of sodium dodecyl sulphate/polyacrylamide-gel electrophoresis, haem, amino acid and terminal-group analysis, and by characterization of haem peptides. Our results confirm and extend those of Kennel *et al.* (1972), and complete the confutation of an earlier report (Dus *et al.*, 1962) of a single peptide to which two haem moieties were attached. There is no similarity between the sequence of the 'di-haem peptide' and the sequence that we believe exists at the haem-attachment site in *Chromatium vinosum* cytochrome *c′* (Fig. 1).

Haem ligand field in cytochromes c′

Cytochromes *c′* have unusual electronic absorption spectra. The spectral and magnetic properties have recently been explained as resulting from a spin 3/2 state with admixtures of spin 5/2 (Maltempo & Moss, 1976), whereas the purely high-spin globins are 5/2 and the low-spin cytochromes are spin 1/2. The ligand-binding properties of cytochromes *c′* are also intermediate between the low-spin proteins, which when native do not combine with small molecules, and the high-spin globins. Cytochromes *c′* will combine with NO and CO (Taniguchi & Kamen, 1963), but the globins will additionally combine with several other substances, such as O_2, Cl^-, SO_3^{2-}, N_3^- and CN^-. The globins only react slowly with oxidizing or reducing agents, but cytochromes *c′* are rapidly reduced by $S_2O_4^{2-}$. Positive charges at the active site of reduction are believed to facilitate the reaction (M. A. Cusanovich, personal communication), and it is possible that a functional requirement for such charges explains why basic residues at positions 122 and 129 (and perhaps 12) (Fig. 1) are conserved in the three cytochromes *c′* of known sequence. As all haem proteins for which a structure is known have histidine as the fifth ligand, and cytochromes *c′* contain a histidine residue in the same position relative to the cysteine residues as in true cytochromes *c*, it is very likely that histidine-125 is the fifth ligand. Cytochromes *c′* lack an invariant methionine residue that could serve as the sixth ligand, as in Type-I (Ambler, 1979) cytochromes *c*, and they also appear to lack a conserved amino acid residue that might provide a weak-field ligand, such as residues containing oxygen (threonine, serine, tyrosine, aspartic acid or glutamic acid).

Support was provided by grants from the Medical Research Council to R. P. A., and from the U.S. National Institutes of Health (GM-18528) and National Science Foundation (B.M.S. 75-3708) to M. D. K.

References

Ambler, R. P. (1973) *Biochem. J.* **135**, 751–758
Ambler, R. P. (1975) *Biochem. J.* **151**, 197–218
Ambler, R. P. (1979) in *Electron Transport in Microorganisms* (Senez, J., LeGall, J. & Peck, H., eds.), C.N.R.S., Paris, in the press
Ambler, R. P. & Wynn, M. (1973) *Biochem. J.* **131**, 485–498
Bartsch, R. G. & Kamen, M. D. (1960) *J. Biol. Chem.* **235**, 825–851
Bartsch, R. G., Coval, M. L. & Kamen, M. D. (1961) *Biochim. Biophys. Acta* **51**, 241–245
Cusanovich, M. A. (1971) *Biochim. Biophys. Acta* **236**, 238–241
Cusanovich, M. A., Tedro, S. M. & Kamen, M. D. (1970) *Arch. Biochem. Biophys.* **141**, 557–570
Dus, K. & Kamen, M. D. (1963) *Biochem. Z.* **338**, 364–375
Dus, K., Bartsch, R. G. & Kamen, M. D. (1962) *J. Biol. Chem.* **237**, 3083–3093
Edman, P. & Begg, G. (1967) *Eur. J. Biochem.* **1**, 80–91
Kennel, S. J., Meyer, T. E., Kamen, M. D. & Bartsch, R. G. (1972) *Proc. Natl. Acad. Sci. U.S.A.* **69**, 3432–3435
Maltempo, M. M. & Moss, T. H. (1976) *Q. Rev. Biophys.* **9**, 181–216
Meyer, T. E., Ambler, R. P., Bartsch, R. G. & Kamen, M. D. (1975) *J. Biol. Chem.* **250**, 8416–8421
Taniguchi, S. & Kamen, M. D. (1963) *Biochim. Biophys. Acta* **74**, 438–455