The amino acid sequence of the dihaem cytochrome c_4 from the bacterium *Azotobacter vinelandii*

R. P. AMBLER,* M. DANIEL,* K. MELIS† and C. D. STOUT†

*Department of Molecular Biology, University of Edinburgh, Edinburgh EH9 3JR, Scotland, U.K., and

†Department of Crystallography, University of Pittsburgh, Pittsburgh, PA 15260, U.S.A.

(Received 22 February 1984/Accepted 27 April 1984)

An amino acid sequence is proposed for the cytochrome c_4 from the bacterium *Azotobacter vinelandii* strain OP. It is a single polypeptide chain of 190 residues, with two sets of haem-attachment cysteine residues at positions 14/17 and 119/122. Proteins with similar sequences are also present in denitrifying pseudomonads. There is similarity in sequence between the two halves of the cytochrome c_4 molecule, and each half also shows similarity to the sequences of certain monohaem cytochromes c isolated from organisms that are not obviously closely related to *A. vinelandii*. Detailed evidence for the amino acid sequence of the protein has been deposited as Supplementary Publication SUP 50125 (17 pages) at the British Library Lending Division, Boston Spa, West Yorkshire LS23 7BQ, U.K., from whom copies are available on prepayment.

The presence of cytochrome c in *Azotobacter* was recognized in the 1930s in several laboratories (Keilin, 1966). In Burris's laboratory, Tissières (1956) purified two distinct cytochromes c from *Azotobacter vinelandii*, and named them cytochromes c_4 and c_5 . These proteins were subsequently crystallized by Neumann & Burris (1959), who showed that (like mitochondrial cytochrome c) they had haem contents of about one per 12000 Da and had acidic isoelectric points (about pH 4.5). Their midpoint oxidoreduction potentials were slightly more positive than that of the mitochondrial protein. Swank & Burris (1969) and Campbell *et al.* (1973) showed that cytochrome c_4 contained two haem groups in a protein of M_r about 21000, and was clearly different in amino acid composition and sequence to the monohaem cytochrome c_5 . The bacterium also produces high concentrations of a third small soluble cytochrome c (Campbell *et al.*, 1973), which possesses an amino acid sequence (Ambler, 1973) that is clearly homologous with that of *Pseudomonas* cytochrome c -551 (Ambler, 1963b).

Although the highest concentrations of cytochromes c seem to occur in *A. vinelandii* cells when they are being forced to fix nitrogen to be able to grow (Swank & Burris, 1969), the cytochromes are not directly involved in this fixation. The physicochemical properties of cytochrome c_4 are not

sufficiently distinctive for its presence in an organism to be demonstrated without purification and molecular characterization (Meyer & Kamen, 1982). However, the cytochrome c -552(II) isolated from a particulate fraction of *Pseudomonas stutzeri* (Kodama & Shidara, 1969) is clearly a cytochrome c_4 , as is a protein from the P -cytochrome-554 fraction of *Pseudomonas aeruginosa* (Horio *et al.*, 1960), as demonstrated by N -terminal sequence analysis (Ambler & Murray, 1973) and other molecular properties.

Despite considerable effort, we have not yet been able to obtain crystals of *A. vinelandii* cytochrome c_4 suitable for high-resolution X-ray-crystallographic analysis. However, Sawyer *et al.* (1981) have been more successful with the protein from *Ps. aeruginosa*, and have reported preliminary crystallographic parameters.

In bacterial cells, cytochrome c_4 appears to be closely associated with membranes, and does not become solubilized as easily as does cytochrome c -551. Purification procedures use extraction with butanol (Tissières, 1956), acetone treatment of a particulate fraction (Kodama & Shidara, 1969) or detergent (Matsushita *et al.*, 1982). Cytochrome c_4 can be extracted from acetone-dried preparations of bacterial cells (Ambler, 1963a), but yields are variable.

In the present paper we propose an amino acid

sequence for *A. vinelandii* cytochrome c_4 , and show that the molecule is made up from two haem-binding domains. There is considerable sequence similarity between these two domains, as well as similarity to the sequences of monohaem cytochromes that have been isolated from strains of *Thiobacillus* and *Paracoccus*.

Experimental

Preparation of cytochrome c_4

Most of the sequence work used protein isolated from *A. vinelandii* strain OP (A.T.C.C. 13705) by the Campbell *et al.* (1973) modification of the method of Swank & Burris (1969). The purity of the protein was assessed spectrally, with values of greater than 1.2 being attained for the ratio of $A_{551(\text{red.})}/A_{270}$, and by sodium dodecyl sulphate/polyacrylamide-gel electrophoresis. The amino acid composition of the purified protein (Table 1) was in satisfactory agreement with that deduced from the sequence (Fig. 1). The material used for the automatic sequenator degradations (Ambler & Murray, 1973) was from strain O (N.C.I.B. 8789, A.T.C.C. 12837), and was isolated from acetone-dried cells by the method of Ambler (1963a). An amino acid analysis of this preparation is also given in Table 1.

Amino acid sequence determination

The amino acid sequence was determined by the general methods that have been used for other bacterial cytochromes c (Ambler & Wynn, 1973; Ambler *et al.*, 1979), and to similar standards. Protein (1–3 μmol) was treated with HgCl_2 in 8M-urea/0.1M-HCl at 37°C for 16 h 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 (5-dimethylaminonaphthalene-1-sulphonyl)/phenyl isothiocyanate method, and by exo- and endopeptidase digestion. The *N*-terminal sequence of the whole protein was investigated with an automatic sequenator (Beckman model 890A) with a fast Quadrol double-cleavage program (Beckman 072172B), identifying the phenylthiohydantoin derivatives by g.l.c. or t.l.c. Amide groups were assigned from peptide electrophoretic mobilities, exopeptidase analysis and direct identification of phenylthiohydantoin derivatives. Peptides containing the –Asp–Pro– bond were cleaved by incubating in 5% (v/v) formic acid at 60°C for 16 h.

Results

The amino acid composition of the protein is shown in Table 1. The evidence for the amino acid sequence proposed is summarized in Fig. 1. Details of the purification, analysis and sequence determination experiments on all these peptides are given in the Supplementary Publication (SUP 50125). The criteria for satisfactory results, and the nature and format of the Supplementary Publication, are given in previous papers (Ambler & Wynn, 1973; Ambler *et al.*, 1979).

The amino acid sequence was deduced by characterizing peptides from four independent digests. These were: (a) 2.3 μmol of (performic acid-)oxidized protein, digested with thermolysin; (b) 2.6 μmol of oxidized protein digested with chymotrypsin; (c) 2.7 μmol of reduced and *S*-aminoethylated protein digested with trypsin; (d) 1.0 μmol of oxidized protein digested with staphylococcal proteinase.

Discussion

Accuracy of proposed sequence

All the peptides are linked into a single sequence by good overlaps, and peptides accounting for the complete postulated sequence were found in each digest (except 'peptide C32'; see below). No peptides that could not be fitted into this sequence were isolated.

The regions on the *C*-terminal sides of the two haem-attachment sites were difficult to sequence, as they did not contain bonds susceptible to the more specific proteinases. Redundant confirmatory sequence information was not obtained for these regions, and the evidence for the assignment of some amide residues is weak. We believe that the sequence of residues 21–23 is –Asp–Gly–Asn– (rather than –Asn–Gly–Asp–) because the amide group is not labile in the way characteristic of –Asn–Gly–sequences. In contrast, peptides containing residues 126–128 (–Asn–Gly–Glu–) did change in electrophoretic mobility in the course of purification, and particularly when treated with 0.1M- NH_3 .

The only evidence given for the amide ascriptions for residues Asp-3 and Gln-8 was derived from automatic sequenator degradations of the whole protein. Evidence from peptide mobilities showed that there was an amide group on one or other of these residues, but the need to determine on which was overlooked until after the suitable peptides had been used for other purposes (fast-atom-bombardment mass spectrometry; Williams *et al.*, 1982).

No attempt was made to characterize peptides formed by CNBr cleavage of the protein, as the

Table 1. *Amino acid composition of Azotobacter vinelandii cytochrome c*₄

Results are shown as mol of residue/mol, and were calculated on the basis that the total of all residues was 188. The samples were hydrolysed at 105°C with 6M-HCl. Sample (3) was hydrolysed for 96 h, the others for 24 h. Sample (1) and (5) were of native protein. The others had had the haem removed before hydrolysis. Samples (2) and (3) were oxidized with performic acid, and sample (4) was aminoethylated with ethyleneimine, before hydrolysis. Samples (1)–(4) were from the protein isolated from strain OP, sample (5) was from the protein from strain O (N.C.I.B. 8789). The 'best value' is the average of the accepted estimates from analyses (1)–(4). Values discarded because the hydrolysis conditions were not optimal for recovery of the amino acid concerned are shown in parentheses. The results reported by Campbell *et al.* (1973) have been recalculated for comparison.

Composition (mol of residue/mol)

Amino acid	Hydrolysis and analysis					'Best value'	Sequence	Campbell <i>et al.</i> (1973)
	(1)	(2)	(3)	(4)	(5)			
Glycine	25.3	24.7	25.8	24.8	23.5	25.2	25	25.4
Alanine	28.8	27.9	28.5	28.3	29.2	28.4	29	30.0
Valine	6.1	5.5	5.8	5.4	5.7	5.7	5	5.2
Leucine	14.4	13.9	14.5	14.4	15.1	14.3	14	13.8
Isoleucine	6.1	5.4	5.9	6.0	6.1	5.9	6	5.1
Serine	10.0	10.5	(8.2)	9.0	9.1	9.8	10	10.2
Threonine	8.5	9.1	(7.9)	9.9	8.0	9.2	9	9.3
Aspartic acid	20.0	19.8	20.6	20.6	19.6	20.3	13	22.8
Asparagine							8	
Glutamic acid	19.1	18.3	18.6	18.7	18.0	18.7	7	19.0
Glutamine							11	
Phenylalanine	4.3	4.0	4.2	4.3	4.3	4.2	4	4.5
Tyrosine	5.4	(4.9)	(4.9)	5.7	5.5	5.6	6	5.3
Tryptophan							0	
Cysteine	(2.7)	3.5*	3.8*	(3.0)†	2.1	3.7	4	2.5
Methionine	5.8	5.5‡	5.9‡	5.7	5.5	5.7	6	5.4
Proline	10.9	10.5	11.0	10.3	10.9	10.7	11	11.4
Lysine	10.6	10.7	11.4	10.3	10.9	10.8	11	9.8
Histidine	4.0	3.7	3.9	3.7	3.6	3.8	4	3.6
Arginine	6.7	7.2	7.0	7.5	8.6	7.1	7	5.2
Total							190	

* As cysteic acid.

† As *S*-aminoethylcysteine.

‡ As methionine sulphone.

distribution of the methionine residues is such that the peptides formed would be likely to be difficult to separate from each other by gel filtration. The seven peptides that should be formed would contain 49, 47, 44, 23, 22, 3 and 2 residues. Campbell *et al.* (1973) isolated and analysed seven peptides after CNBr cleavage of *A. vinelandii* cytochrome *c*₄: two of them can be recognized as fitting the postulated sequence.

The only unexpected proteolytic cleavage observed was that chymotrypsin splits bond 135–136 (–Tyr–Pro–). The predicted peptide 'C32' (Pro–Arg–Leu) was not noticed in the chymotryptic digest, but would have been stained only weakly with ninhydrin on paper, and so could have easily been overlooked. The susceptibility of the bond to chymotrypsin was confirmed by sub-digestion of peptide F15b, although the isatin-staining basic peptide did not give a satisfactory quantitative

amino acid composition, probably because of contamination in the hydrolysis tube. The same sequence –Ala–Tyr–Pro–Arg– occurs in *Leucaena glauca* ferredoxin (Benson & Yasunobu, 1969), and in this case also the –Tyr–Pro– bond appears to be split by chymotrypsin, although the authors do not remark on it. Other plant ferredoxins (e.g. spinach; Matsubara & Sasaki, 1968) lack the arginine residue here, and in these cases the –Tyr–Pro– bond appears to be resistant to chymotryptic hydrolysis.

Features of the amino acid sequence

The polypeptide chain is about twice as long as that found in typical 'Class I' (Ambler, 1982) cytochromes *c*, and there are two haem groups and two normal –Cys–Xaa–Yaa–Cys–His– haem-attachment sequences. When the two haem attachment sites are aligned, considerable sequence similarity can

Fig. 1. Amino acid sequence of *Azotobacter vinelandii* cytochrome c₄. Peptides derived by digestion with trypsin (T) or thermolysin (H) are shown above the sequence, and by digestion with chymotrypsin (C) or staphylococcal proteinase (F) below the sequence. AE indicates peptides formed by tryptic cleavage at S-aminoethiocysteine residues. Peptides from subdigests are labelled with a second letter indicating the secondary method of cleavage (π , mild acid hydrolysis). Full lines indicate quantitative amino acid analyses, substandard if marked *, and particularly bad if marked **. Broken lines indicate qualitative analyses. Under peptide lines, $\rightarrow$ indicates end groups and subsequent residues revealed by phenyl isothiocyanate degradation and identified by the dansyl method, substandard if shown $\rightarrow$. C-Terminal residues of peptides identified as free amino acids after removal of the remainder by phenyl isothiocyanate degradation are shown $\rightarrow$. Peptides marked † were examined by carboxypeptidase A digestion. Under the sequence $\rightarrow$ indicates residues identified as phenylthiohydantoin after degradation of the whole protein in a sequenator, only tentatively identified if marked $\rightarrow$.

	1	2	3	4	5	6	7	8	9
(1A)	AGDAAAGQGGK-AAVCGACHGPDGNSAAPN-FPKLAGQGGERYLLKQMDDIKAGTKPCGAPEGSGRKV-----LEMTGMLDNFSDQDLADLAAYFTSKPTVGAAD---	0	0	0	0	0	0	0	0
(2A)	AGDAAAGQAK-AAVCGACHGpbgbsaapb-fpKLAGQGGERYLLKQMMDIK-----DGKRTV-----LEMTGLLTBlsbzqIADLAAYFASQKMSvgmab---								
(3)	AGDAAAGEDK-IGTCVACHGTGQGLAPI-YPNLTGQSATYLESSIKAYR-----DQGRKG-GNAALMTPMAQGLSDEDIADI AAYYSQEE								
(4)	AGDAAAGKTLVDASCASCHGMQAQGGGM--FPKLAGLTSERIKTTLVAFKSGDTATLKKEGLGGP--NSAINAPNAAGLSEQMDNLSAYIATLK								
(2B)	--pblvzagEALFRGGKIAEGMPACTGCHGpssvgiatagfphLGGQHATYVAKQLTDFR-----EGTRTNgagdtKIMQSI AAKLSNKKDIAAIISSYIQGLH								
(1B)	--PQLVEAGETLYRGGKLDGMPACTGCHSPNGEGNTPAAYPRLSGQHAQYVAKQLTDFR-----EGARTNDGDNMIMRSIAAKLSNKKDIAAIISSYIQGLH	1	1	1	1	1	1	1	1
	0	1	2	3	4	5	6	7	8
	0	0	0	0	0	0	0	0	0

Fig. 2. Alignment of cytochromes c_4 and monohaem cytochrome c sequences

(1A) and (1B), *Azotobacter vinelandii* cytochrome c_4 ; (2A) and (2B), *Pseudomonas aeruginosa* cytochrome c_4 (Ambler, 1982); (3), halotolerant *Paracoccus* sp. N.C.I.B. 8699 cytochrome c -554(540) (Ambler *et al.*, 1984a); (4), *Thiobacillus neapolitanus* cytochrome c -553(547) (Ambler *et al.*, 1984b). The one-letter notation used is that recommended by the IUPAC-IUB Commission on Biochemical Nomenclature [Biochem. J. (1969) 113, 1-4]. In sequences (2A) and (2B) lower-case letters indicate blocks of residues of known composition arranged so as to maximize matching with sequences (1A) and (1B). Vertical lines (|) connect residues that are identical or structurally similar in the sequences.

be recognized between the halves of the molecule (Fig. 2), although three insertions, of one, four and nine amino acids, are needed to maximize matching. A bridge region (residues 90–115) links the matching parts, and contains the only Asp-Pro sequence in the molecule. It should therefore be possible to cleave the cytochrome into two nearly equivalent parts by very mild acid hydrolysis.

An incomplete tentative sequence for the cytochrome c_4 from *Ps. aeruginosa* has been reported (Ambler, 1982; and see Dickerson, 1980). This sequence (Fig. 2) [typographical errors by Ambler (1982) for the *Ps. aeruginosa* sequence caused the omission of the threonine residue corresponding to Thr-159, and showed an extra isoleucine residue after the one corresponding to Ala-172 in the *A. vinelandii* sequence] aligns exactly with that of the *A. vinelandii* protein except that it does not have the nine-residue insertion in the first half of the sequence. There is about 75% sequence identity between the two proteins.

Each half of cytochrome c_4 shows sequence similarity to the monohaem cytochrome c of Class IC (Ambler, 1982), which include the soluble 'cytochrome f ' from eukaryotic algae and cyanobacteria (Ambler & Bartsch, 1975; see Wood, 1977, for nomenclature clarification), and proteins from various bacteria with nitrate or sulphur-compound metabolisms. We expect that when the three-dimensional structure of a cytochrome c_4 is elucidated, each half of the sequence will be found to take up the 'cytochrome fold' (Dickerson, 1980). Of the six different Class I cytochromes c for which structures are available, we suggest that cytochrome c_4 segments may most resemble the cytochrome c -554 from the cyanobacterium *Anacystis nidulans* (Ludwig *et al.*, 1982) and *Chlorobium* cytochrome c -555 (Korszun & Salemme, 1977). The most similar monohaem protein so far found is the cytochrome c -554(548) from the halotolerant *Paracoccus* sp. N.C.I.B. 8699 (=A.T.C.C. 12084, I.A.M. 203) (Hori, 1961a,b; Ambler *et al.*, 1984a), a protein that forms very stable non-covalent dimers. Another protein that matches well is the cytochrome c -554(547) from *Thiobacillus neapolitanus* (N.C.I.B. 8539) (Ambler *et al.*, 1984b). It is interesting that sequence revision shows that the cytochrome c -553 from *Desulfovibrio vulgaris* can now be accommodated in this class (Nakano *et al.*, 1983).

In the alignment shown in Fig. 2, each of the sequences contain methionine at the positions corresponding to residues 66 and 167 in the *A. vinelandii* cytochrome c_4 sequence, and these residues are likely to form the sixth ligands to the iron in the haem moieties. Detailed three-dimensional electron density maps are not yet available for any of these cytochromes, but crystallographic studies of

the *Ps. aeruginosa* cytochrome c_4 are well advanced (Sawyer *et al.*, 1981).

This work was supported by grants to R. P. A. from the Medical Research Council and Science and Engineering Research Council, and to C. D. S. from the National Institutes of Health (GM-25672).

References

- Ambler, R. P. (1963a) *Biochem. J.* **89**, 341–349
- Ambler, R. P. (1963b) *Biochem. J.* **89**, 349–378
- Ambler, R. P. (1973) *Syst. Zool.* **22**, 554–565
- Ambler, R. P. (1982) in *From Cyclotrons to Cytochromes* (Robinson, A. R. & Kaplan, N. O., eds.), pp. 263–280, Academic Press, New York
- Ambler, R. P. & Bartsch, R. G. (1975) *Nature (London)* **253**, 285–288
- Ambler, R. P. & Murray, S. (1973) *Biochem. Soc. Trans.* **1**, 162–164
- 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) *Biochem. J.* **177**, 819–823
- Ambler, R. P., McLellan, L., Meyer, T. E. & Kamen, M. D. (1984a) *Biochem. J.* in the press
- Ambler, R. P., Meyer, T. E., Trudinger, P. A. & Kamen, M. D. (1984b) *Biochem. J.* in the press
- Benson, A. M. & Yasunobu, K. T. (1969) *J. Biol. Chem.* **244**, 955–963
- Campbell, W. H., Orme-Johnson, W. H. & Burris, R. H. (1973) *Biochem. J.* **135**, 617–630
- Dickerson, R. E. (1980) *Sci. Am.* **242**, 98–110
- Hori, K. (1961a) *J. Biochem. (Tokyo)* **50**, 440–449
- Hori, K. (1961b) *J. Biochem. (Tokyo)* **50**, 481–485
- Horio, T., Higashi, T., Sasagawa, M., Kusai, K., Nakai, M. & Okunuki, K. (1960) *Biochem. J.* **77**, 194–201
- Keilin, D. (1966) *The History of Cell Respiration and Cytochrome*, Cambridge University Press, London
- Kodama, T. & Shidara, S. (1969) *J. Biochem. (Tokyo)* **65**, 351–360
- Korszun, Z. R. & Salemme, F. R. (1977) *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5244–5247
- Ludwig, M. L., Dickerson, R. E. & Takano, T. (1982) in *Electron Transport and Oxygen Utilization* (Ho, C., ed.), pp. 27–32, Elsevier, Amsterdam
- Matsubara, H. & Sasaki, R. M. (1968) *J. Biol. Chem.* **243**, 1732–1757
- Matsushita, K., Shinagawa, E., Adachi, O. & Ameyama, M. (1982) *J. Biochem. (Tokyo)* **92**, 1607–1613
- Meyer, T. E. & Kamen, M. D. (1982) *Adv. Protein Chem.* **35**, 105–212
- Nakano, K., Kikumoto, Y. & Yagi, T. (1983) *J. Biol. Chem.* **258**, 12409–12412
- Neumann, N. P. & Burris, R. H. (1959) *J. Biol. Chem.* **234**, 3286–3290
- Sawyer, L., Jones, C. L., Damas, A. M., Harding, M. M., Gould, R. O. & Ambler, R. P. (1981) *J. Mol. Biol.* **153**, 831–835
- Swank, R. T. & Burris, R. H. (1969) *Biochim. Biophys. Acta* **180**, 473–489
- Tissières, A. (1956) *Biochem. J.* **64**, 582–588
- Williams, D. H., Bradley, C. V., Santikarn, S. & Bojesen, G. (1982) *Biochem. J.* **201**, 105–117
- Wood, P. M. (1977) *Eur. J. Biochem.* **72**, 605–612