The amino acid sequence of the cytochrome *c*-554(547) from the chemolithotrophic bacterium *Thiobacillus neapolitanus*

R. P. AMBLER,* T. E. MEYER,† P. A. TRUDINGER‡ and M. D. KAMEN§

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

†Department of Biochemistry, University of Arizona, Tucson, AR 85721, U.S.A., ‡Baas Becking Geobiological Laboratory, P.O. Box 378, Canberra, A.C.T. 2601, Australia, and §1390 Plaza Pacifica, Santa Barbara, CA 93108, U.S.A.

(Received 14 January 1985/14 February 1985; accepted 21 February 1985)

An amino acid sequence is proposed for the cytochrome *c*-554(547) from the bacterium *Thiobacillus neapolitanus* N.C.I.B. 8539). It consists of a polypeptide chain of 91 residues, with a pair of haem-attachment cysteine residues at positions 15 and 18. There is similarity in sequence with each of the halves of the sequence of the dihaem cytochromes *c*₄ and with a cytochrome *c*-554(548) from a halophilic strain of *Paracoccus*. Detailed evidence for the amino acid sequence of the protein has been deposited as Supplementary Publication SUP 50127 (11 pages) at the British Library (Lending Division), Boston Spa, Wetherby, West Yorkshire LS23 7BQ, U.K., from whom copies can be obtained on the terms indicated in *Biochem. J.* (1985) **225**, 5.

Thiobacillus neapolitanus was first isolated (as *Thiobacillus* X; Parker, 1947; Parker & Prisk, 1953) as one of the organisms responsible for the early stages of bacterial concrete corrosion, in which the alkaline state (pH 11–12) of the fresh concrete surface is lowered to a just-acidic value (pH 6.5). From this pH, the H₂SO₄-forming strains of '*Thiobacillus concretivorus*' (now incorporated into *Thiobacillus thiooxidans*) take over, and the environment becomes extremely acidic. *T. neapolitanus* is an aerobic, strictly autotrophic, organism that oxidizes sulphur compounds, such as thiosulphate, to sulphate. It has a pH optimum of about 6.5, but tolerates pH 3–8.5 (Vishniac, 1974). The organism has long been known to produce large amounts of soluble cytochromes *c* (Trudinger, 1961), and the cytochrome complement has recently been reinvestigated by modern methods (Trudinger *et al.*, 1984).

T. neapolitanus under normal growth conditions produces two predominating cytochromes *c*. Cytochrome *c*-554(547) is small (*M*_r < 12000) and monomeric, whereas cytochrome *c*-549 was isolated as aggregates of a 26000-*M*_r subunit (Trudinger

et al., 1984). Cytochrome *c*-554(547) has a visible absorption spectrum in which the α -peak is split into two well-resolved peaks at room temperature, the maxima being about the same as that of the β -peak. Most cytochromes *c* show this splitting of the α -band at liquid-nitrogen temperature. The cytochrome *c*-554(547) has a slightly acidic pI (Trudinger *et al.*, 1984) and is unreactive in the bovine cytochrome oxidase and reductase enzyme assays (Errede, 1976).

In the present paper we report the amino acid sequence of the cytochrome *c*-554(547). The sequence shows some similarity to each domain of the dihaem cytochrome *c*₄ (Ambler *et al.*, 1984) and to the cytochrome *c*-554(548) of a halophilic *Paracoccus* strain (N.C.I.B. 8669) (Hori, 1961*a,b*; R. P. Ambler, L. McLellan, T. E. Meyer & M. D. Kamen, unpublished work).

Experimental

Preparation of cytochrome *c*-554(547)

The cytochrome was purified as described by Trudinger *et al.* (1984). The yield from each of two 200g-wet-weight batches of cells was 6 μ mol. The final preparation had no 280nm absorption peak, and the ratio of *A*_{554(reduced)}/*A*₂₈₀ was higher than 1.2. The purity of the preparation is discussed below.

Abbreviations used: dansyl, 5-dimethylaminonaphthalene-1-sulphonyl; PTH, phenylthiohydantoin; DMBA, *NN*-dimethylbenzylamine; SDS, sodium dodecyl sulphate.

Amino-acid-sequence determination

The amino acid sequence was investigated 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 (2 µmol) was treated with HgCl₂ in 8M-urea/0.1 M-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. In other experiments the native protein was cloven with CNBr. 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 phenylisothiocyanate method, and by exo- and endo-peptidase digestion. The *N*-terminal sequence of the whole native protein was investi-

gated with an automatic sequenator (Beckman model 890A) with a DMBA program (Beckman 102473), the PTH derivatives being identified by g.l.c. and t.l.c. Amide groups were assigned from peptide electrophoretic mobilities and exopeptidase analysis. Pyroglutamate (pyrrolidonecarboxylic acid) aminopeptidase (EC 3.4.11.8; Boehringer Mannheim) was used to unblock one peptide, which was treated with the enzyme (0.025 Boehringer units/160 nmol) in 0.1% (w/v) dithiothreitol/0.2 M-acetic acid adjusted to pH 8.5 with NH₃ at 37°C for 5 h, and the product isolated by paper electrophoresis.

Results

The amino acid composition of the protein preparation is shown in Table 1. The evidence for

Table 1. *Amino acid composition of Thiobacillus neapolitanus cytochrome c-554(547)*

Results were calculated on the basis that the total of all residues was 91. The samples were hydrolysed at 105°C with 6M-HCl. Sample (3) was hydrolysed for 96 h, the others for 24 h. Samples (1) and (5) were of native protein. The others had had the haem removed before hydrolysis. Sample (4) was oxidized with performic acid before hydrolysis. Sample (5) was from a batch of protein purified independently from that used for samples (1)–(4). The 'Best value' is the average of the accepted estimates. Values discarded because the hydrolysis conditions were not optimal for recovery of the amino acid concerned or because peak resolution was inadequate because of instrument malfunction are shown in parentheses.

Composition (residues/molecule)									
Amino acid	Sample	...	Hydrolysis and analysis					'Best value'	From sequence
			(1)	(2)	(3)	(4)	(5)		
Glycine			11.8	11.9	12.0	11.9	12.2	12.0	11
Alanine			17.5	18.4	18.8	18.3	18.2	18.2	16
Valine			(1.5)*	0.9	1.0	0.9	0.9	0.9	1
Leucine			8.7	9.4	9.6	8.9	9.1	9.1	9
Isoleucine			2.7	2.8	3.1	2.6	3.0	3.1	3
Serine			6.3	5.6	(5.1)	(6.9)*	5.7	5.9	7
Threonine			6.9	7.3	6.8	(7.6)*	6.6	6.9	7
Aspartic acid			8.3	8.6	8.4	(8.2)*	8.3	8.4	5
Asparagine									3
Glutamic acid			7.1	6.9	7.0	(6.6)*	6.7	6.9	3
Glutamine									4
Phenylalanine			2.1	2.1	2.1	2.1	2.1	2.1	2
Tyrosine			1.7	1.9	1.9	(1.7)	1.8	1.8	2
Tryptophan								<0.04†	0
Cysteine				(1.0)	(1.3)	2.0‡		2.0	2
Methionine			3.6	3.7	3.8	(3.4)*§	3.8	3.7	4
Proline			3.4	3.0	3.1	2.8	3.3	3.1	3
Lysine			7.1	6.6	*	6.6	7.1	6.9	7
Histidine			1.2	1.2	*	1.2	1.1	1.2	1
Arginine			0.9	0.9	*	0.9	0.9	0.9	1
Total									91

*Values discarded or not calculated because of instrument malfunction.

†Estimated from analysis of a sample of native protein after hydrolysis with 3M-mercaptoethanesulphonic acid for 24 h at 205°C.

‡As cysteine acid.

§As methionine sulphone.

Fig. 1. Amino acid sequence of *Thiobacillus neapolitanus* cytochrome *c*-554(547)

Peptides derived by digestion with trypsin (T) are shown above the sequence, and by digestion with chymotrypsin (C) or cleavage with CNBr (X) below the sequence. Peptides from subdigests are labelled with a second letter indicating the secondary method of cleavage (F, staphylococcal proteinase; γ, pyroglutamate aminopeptidase). Continuous lines indicate quantitative amino acid analyses, substandard if marked *, and particularly bad if marked **. Broken lines indicate low-yield peptides that were identified but not further characterized. Under peptide lines, → indicates end groups and subsequent residues revealed by phenylisothiocyanate degradation and identified by the dansyl method, substandard if shown - γ. C-Terminal residues of peptides isolated as free amino acids after removal of the remainder by phenylisothiocyanate degradation are shown ▸. Peptides marked † were examined by carboxypeptidase A digestion, and those marked ‡ by digestion with aminopeptidase M. Under the sequence → indicates residues identified as phenylthiohydantoin after degradation of the whole protein in a sequenator, only tentatively identified if marked - →.

the proposed amino acid sequence is summarized in Fig. 1. Details of the purification, analysis and sequence experiments on the peptides are given in Supplementary Publication SUP 50127. 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 separate digests. These were: (a) 2.0 mol digested with trypsin; (b) 2.4 μ mol digested with chymotrypsin; (c) 2.5 μ mol of native protein cloven with CNBr; and (d) 1.9 μ mol of native protein cloven with CNBr.

Discussion

Purity of initial protein preparation

Although the preparation had been purified until rechromatography did not improve the spectral ratios, which were those of a pure cytochrome *c* that did not contain any tryptophan, the sequence experiments suggested that the preparation contained another polypeptide component present at about a 10% molar level.

This contaminant became manifest when low levels of peptides that could not be accounted for in the proposed sequence (Fig. 1) were detected in the chymotryptic digest. The chymotryptic digestion has resulted in several partial cleavages, with the result that the mixture was more complex than expected and yields of some individual peptides were low. Some fractions needed more steps of fractionation than usual, and some of the high-yield fragments from the putative contaminant were isolated at levels (0.03 mol/mol of cytochrome) not much less than those of poor-yield peptides from the cytochrome. In contrast, in the tryptic digest, bond cleavage was very specific, so

the yields of all the cytochrome peptides were good, and the occurrence of low levels of one or two spurious peptides was only recognized in retrospect. One incompatible peptide was isolated at a 3% level from the CNBr digests. Its composition included arginine, histidine and tyrosine, amino acids occurring at only low levels in the cytochrome (Table 1). The incompatible peptides that were found and partly characterized are listed in the supplementary publication.

The amino acid composition of the protein preparation (Table 1) agrees with that of the deduced sequence for the amino acids present as small numbers of residues, but has 14% excess alanine, 9% excess glycine and 16% too little serine. These discrepancies would be largely explained were residue 14 or 17 (in the difficult-to-sequence haem-attachment region) alanine rather than serine, but the peptide-analysis values make this explanation unlikely. In view of the peptide evidence for a contaminant polypeptide, the agreement shown in Table 1 is surprisingly good.

SDS/polyacrylamide-gel electrophoresis was not a satisfactory way of assessing the homogeneity of the cytochrome preparation, since with the Amido Black staining system used, the protein washed out of the gel. It could be seen migrating as a faint red band that did stain with Amido Black but was washed out in the acid methanol.

No contaminant sequence was detected in the sequenator degradation of the cytochrome, but the system used for detecting the PTH derivatives was not quantitative, and low levels of another sequence would have been missed.

Difficulties in the sequence determination

The composition of residues 22–26 could be reliably deduced from analyses of peptides, but the sequence was not accessible by degradation of

Fig. 2. Alignment of *Thiobacillus neapolitanus* cytochrome *c*-554(547) sequence with sequences of other bacterial cytochromes *c* (from Ambler *et al.*, 1984)

(1A) and (1B), *Azotobacter vinelandii* cytochrome *c*₄; (2A) and (2B), *Pseudomonas aeruginosa* cytochrome *c*₄; (3), halophilic *Paracoccus* sp. N.C.I.B. 8669 cytochrome *c*-554(540); (4), *Thiobacillus neapolitanus* cytochrome *c*-554(547). 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.

tryptic or chymotryptic peptides. The region was isolated as a CNBr peptide that did not react with ninhydrin. The behaviour of the peptide during purification by gel filtration and paper electrophoresis was predicted from its deduced composition, and it was finally detected by acid hydrolysis of samples of the likely fractions. After removal of one 'glutamic acid' residue with pyroglutamate aminopeptidase, the residue did react with ninhydrin, and was analysed and sequenced by the standard methods.

The assignment of amide groups to the five possible sites in the region of residues 77–82 requires the use of exopeptidases and careful interpretation of electrophoretic-mobility results.

Structural affinities of T. neapolitanus cytochrome c-554(547)

The sequence shown in Fig. 1 has considerable similarity (see Fig. 2) to that of the cytochrome *c*-554(548) from the halophilic *Paracoccus* sp. N.C.I.B. 8669 (R. P. Ambler, L. McLennan, T. E. Meyer & M. D. Kamen, unpublished work), which also has a similar split α -band spectrum. The sequence also resembles those of both domains of cytochrome *c*₄ (present Fig. 2; Ambler *et al.*, 1984). The *N*-terminal seven residues of sequence are identical in all these cytochromes, whereas similar, though not identical, sequences occur at the *N*-terminus of some cytochromes *c* of known tertiary structure (Dickerson, 1980), where they form the *N*-terminal α -helices.

There is no obvious metabolic similarity between *T. neapolitanus* and the organisms that produce the cytochromes of similar sequences.

We thank Margaret Daniel and Lesley McLellan for skilled and enthusiastic assistance. The work was supported by grants from the Medical Research Council and the Science and Engineering Research Council to R.P.A., and from the National Institutes of Health (GM 18528) and the National Science Foundation (PCM 83-11944) to M. D. K.

References

- 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., Daniel, M., Melis, K. & Stout, C. D. (1984) *Biochem. J.* **222**, 217–227
- Dickerson, R. E. (1980) *Sci. Am.* **242**, 98–110
- Errede, B. J. (1976) Ph.D. Thesis, University of California at San Diego
- Hori, K. (1961a) *J. Biochem. (Tokyo)* **50**, 440–449
- Hori, K. (1961b) *J. Biochem. (Tokyo)* **50**, 481–485
- Parker, C. D. (1947) *Nature (London)* **159**, 439–440
- Parker, C. D. & Prisk, J. (1953) *J. Gen. Microbiol.* **8**, 344–364
- Trudinger, P. A. (1961) *Biochem. J.* **78**, 673–680
- Trudinger, P. A., Meyer, T. E., Bartsch, R. G. & Kamen, M. D. (1984) *Arch. Microbiol.* in the press
- Vishniac, W. V. (1974) in *Bergey's Manual of Determinative Bacteriology*, 8th edn. (Buchanan, R. E. & Gibbons, N. E., eds.), pp. 456–461, Williams and Wilkins, Baltimore