A reinvestigation of the covalent structure of *Pseudomonas aeruginosa* cytochrome c peroxidase

Bart Samyn[a], Kathleen Van Craenenbroeck[a], Lina De Smet[a], Isabel Vandenberghe[a], Graham Pettigrew[b], Jozef Van Beeumen[a,*]

[a]*Department of Biochemistry, Physiology and Microbiology, Laboratory of Protein Structure and Function, University of Gent, Ledeganckstraat 35, 9000 Gent, Belgium*

[b]*Department of Preclinical Veterinary Sciences, Royal (Dick) School of Veterinary Studies, University of Edinburgh, Summerhall, Edinburgh EH9 1QH, UK*

Received 27 October 1995

Abstract The amino acid sequence of cytochrome c peroxidase from *Pseudomonas aeruginosa* has been determined using classical chemical degradation techniques combined with accurate mass analysis of all the generated peptides. The sequence obtained is composed of 346 amino acids and confirms the recently published cDNA-derived sequence except at one position [Ridout et al. (1995) FEBS Lett. 365, 152–154]. Based on this sequence, we propose a new model for the binding of the peroxide and the cytochrome electron donor to CCP which is in essence the reverse of the one proposed by Ellfolk et al. [Biochim. Biophys. Acta 1080 (1991) 175–178].

Key words: Cytochrome c peroxidase; *Pseudomonas aeruginosa*; Mass analysis; Primary structure

1. Introduction

Cytochrome c peroxidase (CCP) from *Pseudomonas aeruginosa* (Pa) is the only bacterial CCP that has so far been sequenced. Rönnberg et al. [1] originally reported that the polypeptide is 302 amino acids long containing two typical haem c-binding sites (Cys-X-Y-Cys-His), and that it had six positions where the presence of two residues betrayed micro heterogeneity. In a subsequent paper, Ellfolk et al. [2] corrected this sequence by reducing the number of heterogeneous positions to two. The authors proposed that the first half of the polypeptide chain contains the low potential haem-binding region that reacts with the hydrogen peroxide, and that the second half contains the high potential haem-binding region that interacts with the reduced electron donor protein of CCP. Carrying out high precision mass analysis of Pa CCP, we recently noticed that its molecular mass is some 2 kDa higher than the calculated mass of the amino acid sequence published by Ellfolk et al. [2]. We now report the complete amino acid sequence of PaCCP, including the proof of the covalent linkage of the two haem groups, based on the evidence obtained from Edman degradation in combination with mass analysis of the sequenced peptides. The data correct the original sequence at several positions including, as a major point, the presence of an additional section of nineteen residues. Our results were presented at the Protein Society Meeting held in Davos, Switzerland, from 28 May to 1 June 1995. Shortly thereafter, Ridout et al. [3] reported the nucleotide sequence of PaCCP which is in agreement with our sequence except at one position.

*Corresponding author. Fax: (32) (9) 264 53 38.
E-mail: jozef.vanbeeumen@rug.ac.be

Abbreviations: RPLC, reversed-phase liquid chromatography; ESMS, electrospray mass spectrometry; MALDI, matrix-assisted laser desorption mass spectrometry.

2. Methods and materials

2.1. Enzymatic and chemical cleavages

About 1 nmol holo CCP was digested with endoproteinase Lys-C (Boehringer, Mannheim, Germany) (50 mM Tris-HCl, pH 8.0, E/S = 1/40, 7.5 h, 37°C). A second cleavage was performed by partial acid hydrolysis on about 2 nmol holo CCP for 3 h at 106°C (this cleavage was performed twice). A final digest was performed with endoproteinase Glu-C (Boehringer) on 6 nmol apo-CCP (100 mM NH_4HCO_3, pH 7.6, E/S = 1/40. One Glu-C peptide was subjected to a subdigest with endoproteinase Arg-C (Boehringer). The subdigest was carried out according the instructions given by the company.

2.2. Peptide purification

Half of the peptide material resulting from the Lys-C-digest and from the partial acid hydrolysis was subjected to RPLC analysis on a Sephasil C_{18} column (2.1 × 100 mm) (Pharmacia, Uppsala, Sweden) using a gradient of solvent A (0.05% TFA in water) and solvent B (0.04% TFA in 80% ACN). The peptides from the Glu-C-digest were purified using a fast desalting column (PC 3.2 × 100 mm) (Pharmacia) eluted with 5% formic acid. The first peak, containing two larger peptides, was subjected to RPLC on a TMS column (BioRad, Nazareth Eke, Belgium); the smaller peptides in the second gel filtration fraction were separated on a mRPC C_2/C_{18} column (2.1 × 100 mm, Pharmacia). Peptides from the Arg-C subdigest were separated on a Sephasil C_8 column (2.1 × 100 mm) (Pharmacia) and eluted with a linear gradient of water/ACN. Peptides were detected by UV absorbance at 220 and 280 nm. The separation of the second half of the peptide mixtures from the Lys-C digest and the partial acid hydrolysis was performed by on-line LCMS. The system was described in detail by Klarskov et al. [5].

2.3. Sequence analysis

Sequence analysis was performed on the models 477A and 476A pulsed liquid-phase sequenators with on-line PTH analysis on a 120A analyser (Applied Biosystems, Foster City, CA). Sequencing reagents and solvents were obtained from the same firm.

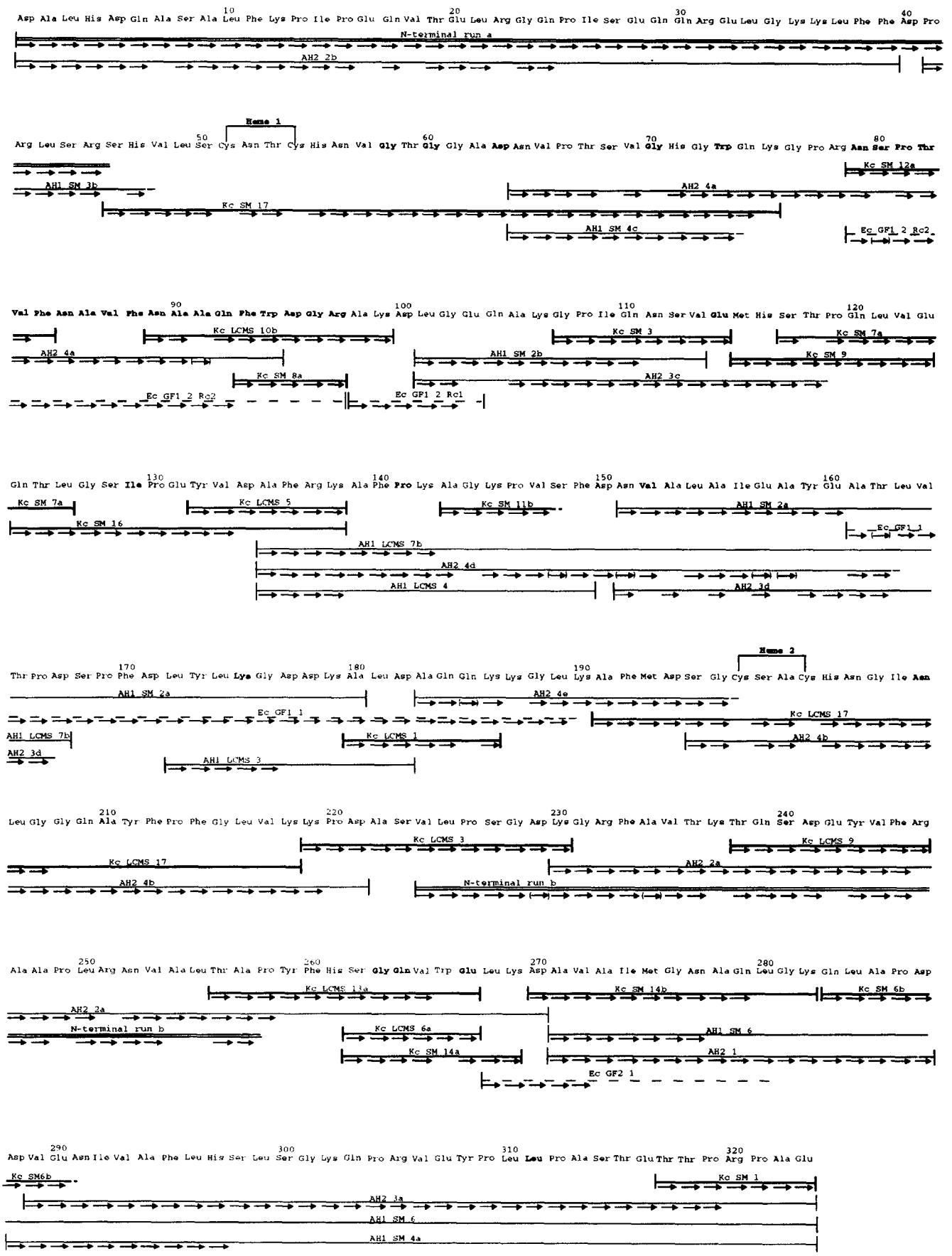
2.4. Mass spectrometry

All LCMS analyses were performed on a BIO-Q triple-quadrupole mass spectrometer equipped with a pneumatically assisted electrospray source (Fisons, Altrincham, UK) operating in the positive ion mode. Some peptides were also subjected to MALDI-TOF analysis. The samples were analysed on a VG Tofspec (Fisons, VG Analytical, Wythenshawe, UK) equipped with a Nitrogen laser [6].

3. Results

The study started by analysing the mass of native PaCCP

SSDI 0014-5793(95)01326-1


Asp Ala Leu His Asp Gln Ala Ser Ala Leu Phe Lys Pro Ile Pro Glu Gln Val Thr Glu Leu Arg Gly Gln Pro Ile Ser Glu Gln Arg Glu Leu Gly Lys Lys Leu Phe Phe Asp Pro
N-terminal run a
AH2 2b
Heme 1
Arg Leu Ser Arg Ser His Val Leu Ser Cys Asn Thr Cys His Asn Val Gly Thr Gly Gly Ala Asp Asn Val Pro Thr Ser Val Gly His Gly Trp Gln Lys Gly Pro Arg Asn Ser Pro Thr
AH1 SM 3b
Kc SM 17
AH2 4a
Kc SM 12a
AH1 SM 4c
Ec GF1 2 Rc2
Val Phe Asn Ala Val Phe Asn Ala Ala Gln Phe Trp Asp Gly Arg Ala Lys Asp Leu Gly Glu Gln Ala Lys Gly Pro Ile Gln Asn Ser Val Glu Met His Ser Thr Pro Gln Leu Val Glu
Kc LCMS 10b
AH2 4a
Kc SM 8a
Ec GF1 2 Rc2
Ec GF1 2 Rc1
Kc SM 3
AH1 SM 2b
AH2 3c
Kc SM 7a
Kc SM 9
Gln Thr Leu Gly Ser Ile Pro Glu Tyr Val Asp Ala Phe Arg Lys Ala Phe Pro Lys Ala Gly Lys Pro Val Ser Phe Asp Asn Val Ala Leu Ala Ile Glu Ala Tyr Glu Ala Thr Leu Val
Kc SM 7a
Kc SM 16
Kc LCMS 5
Kc SM 11b
AH1 SM 2a
Ec GF1 1
AH1 LCMS 7b
AH2 4d
AH1 LCMS 4
AH2 3d
Heme 2
Thr Pro Asp Ser Pro Phe Asp Leu Tyr Leu Lys Gly Asp Asp Lys Ala Leu Asp Ala Gln Gln Lys Lys Gly Leu Lys Ala Phe Met Asp Ser Gly Cys Ser Ala Cys His Asn Gly Ile Asn
AH1 SM 2a
Ec GF1 1
AH1 LCMS 7b
AH2 3d
AH1 LCMS 3
Kc LCMS 1
AH2 4e
Kc LCMS 17
AH2 4b
Leu Gly Gly Gln Ala Tyr Phe Pro Phe Gly Leu Val Lys Lys Pro Asp Ala Ser Val Leu Pro Ser Gly Asp Lys Gly Arg Phe Ala Val Thr Lys Thr Gln Ser Asp Glu Tyr Val Phe Arg
Kc LCMS 3
Kc LCMS 17
AH2 4b
Kc LCMS 9
AH2 2a
N-terminal run b
Ala Ala Pro Leu Arg Asn Val Ala Leu Thr Ala Pro Tyr Phe His Ser Gly Gln Val Trp Glu Leu Lys Asp Ala Val Ala Ile Met Gly Asn Ala Gln Leu Gly Lys Gln Leu Ala Pro Asp
Kc LCMS 13a
AH2 2a
N-terminal run b
Kc LCMS 6a
Kc SM 14a
Kc SM 14b
Kc SM 6b
AH1 SM 6
AH2 1
Ec GF2 1
Asp Val Glu Asn Ile Val Ala Phe Leu His Ser Leu Ser Gly Lys Gln Pro Arg Val Glu Tyr Pro Leu Leu Pro Ala Ser Thr Glu Thr Thr Pro Arg Pro Ala Glu
Kc SM6b
AH2 3a
AH1 SM 6
AH1 SM 4a
Kc SM 1

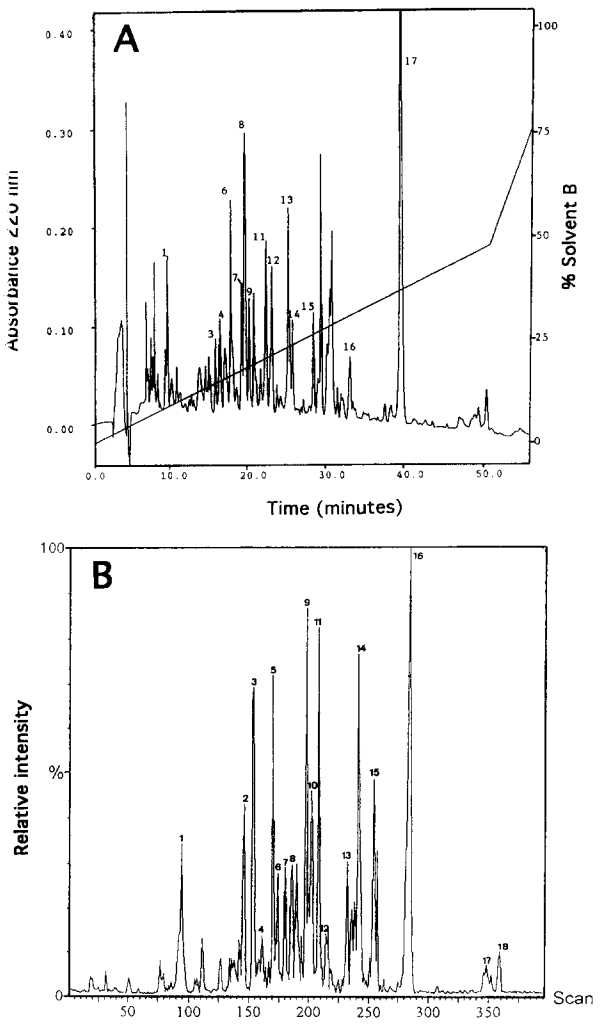


Fig. 2. (A) RPLC separation (SMART) of the Lys-C peptides of native PaCCP. The gradient profile of solvent B is indicated (- - - - -). Numbers designate the resulting peptides. (B) On-line LCMS analysis of the Lys-C peptides of native PaCCP. The *y*-axis represents the base peak intensity of the measured ion current (BPI), the *x*-axis represents the number of mass spectra acquired during the chromatographic analysis. Numbers designate the resulting peptides.

using both ESMS and LDMS (36,251.4 Da and 36,215.3 Da, respectively, data not shown). These masses indicated that the polypeptide is actually some 20 residues longer than the sequence published by Rönnberg et al. [1]. Fig. 1 shows the complete sequence of PaCCP. N-terminal sequence analysis of the native protein revealed the identity of the first 45 amino acids and the presence of a minor fragment (20%) starting at position Val-224, apparently originating from a Ser–Val cleavage. Subsequently, the holoprotein was cleaved with Lys-C endoproteinase as well as by partial acid hydrolysis (5% formic acid). Both peptide mixtures were separated by RPLC as given under section 2 (Fig. 2A gives the result for the Lys-C digest) as well as by mass analysis of the eluted peptides by LCMS (Fig. 2B). Peptides Kc SM 12a and Kc LCMS 10b appeared to contain sequence information that was not presented previously [1,2]. Fraction AH2 4, through one of its component peptides (4a), linked the peptides Kc SM 12a and Kc LCMS 10b. Mass analysis of this peptide yielded a value of 3415.6 Da corresponding with the calculated mass 3414.8 Da (Asn-64–Trp-94) and confirming the presence of the peptide section Asn-79–Arg-97. In order to prove the remaining overlaps, a Glu-C endoproteinase digest was performed on the apoprotein. Two larger fragments of, respectively, 82 (Ala-161–Glu-242) and 71 (Leu-33–Glu-103) amino acids were separated by gel filtration analysis and further purified by reversed-phase HPLC. The measured mass of the 82 residue peptide (8878.4 Da) was approximately 200 Da higher than the calculated value (8679.8 Da). This extra mass is due to a Hg-atom bound to one of the cysteine residues 197 or 200, originating from the method of haem removal [7]. The second large peptide fraction (Leu-33–Glu-103) was subdigested with endoproteinase Arg-C and yielded the peptides Rc2 and Rc1 which proved the overlap for the Kc and AH peptides found in the region Asn-90–Glu-103.

4. Discussion

The molecular mass calculated from the proposed sequence supplemented with two covalently bound haem groups is 36,234.3 Da. The difference of 17.1 Da with the ESMS value of 36,251.4 Da betrays the presence of an oxidized residue, most likely of a methionine. The combination of the mass data (Tables 1 and 2) with the sequence data of the individual peptides,

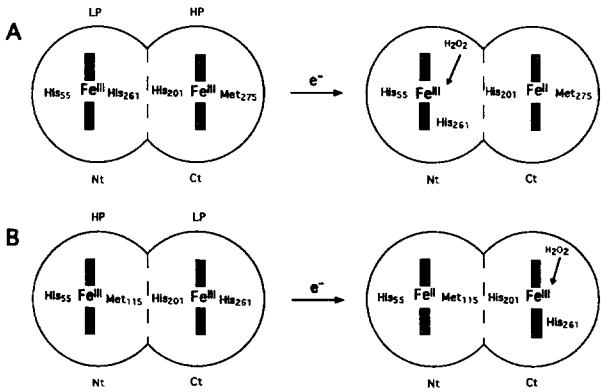


Fig. 3. Models for the mechanism of action of PaCCP. (A) Model proposed by Ellfolk et al. [2]; (B) model deduced from the revised sequence (this work). HP and LP indicate high and low potential haems, respectively. Nt and Ct represents the first and second half of the protein. The peroxide-binding site is indicated by the reagent H_2O_2.

←
Fig. 1. The complete amino acid sequence of CCP from *P. aeruginosa*. Peptides obtained by digestion of the native protein with endoproteinase Lys-C separated by RPLC on the SMART system or by LCMS are indicated as Kc SM and Kc LCMS, respectively. Similarly, peptides obtained after partial acid hydrolysis of the nativew protein are shown as AH1 SM or AH1 LCMS. Peptides AH2 resulted from the second partial acid hydrolysis were separated by RPLC analysis. Peptides of the apoprotein digested with endoproteinase Glu-C and separated by gel filtration and RPLC analysis have the label Ec GF. The peptide Ec GF1 fraction TMS 2 was subdigested with endoproteinase Arg-C and the resulting peptides are shown as Ec GF1 2 Rc. Only those peptides are shown which were considered to be essential for deducing the sequence. Residues indicated in bold are corrections and/or insertions with respect to the first sequence published [1].

Table 1
Mass spectrometric results of Lys-C peptides by MALDI analysis after RPLC separation (A) and by ESMS analysis after LCMS separation (B)

Sequence position	Calculated mass	(A) Fractions from SMART		(B) Fractions from LCMS	
		Fraction	MALDI	Fraction	ESMS
Ala-180–Lys-186	772.8			1	772.4
Phe-233–Lys-237	564.7			2	565.6
Lys-219–Lys-230	1213.4			3	1212.9
Leu-310–Glu-316	729.8			4	730.9
Tyr-132–Lys-138	898.0			5	898.0
Ser-262–Glu-267	704.7			6	705.6
Thr-317–Glu-323	770.8	1	771.7		
Gly-107–Glu-114	842.9	3	843.8		
Leu-154–Thr-165	1293.5	4	1295.8		
Ser-117–Leu-126	1115.2	7 a	1114.4		
Thr-256–His-261	735.8	b	737.2		
Phe-93–Arg-97	679.7	8 a	679.9		
Gln-17–Glu-28	1356.5	b	1357.4		
Met-115–Glu-123	1041.2	9	–	7	1040.9
Asp-171–Lys-175	650.8			8	651.8
Thr-238–Arg-246	1144.2	11 a	1143.9	9	1144.6
Ala-143–		b	–		
Asn-79–Phe-84	663.7	12 a	–	10 a	664.8
Asn-89–Lys-99	1263.4	b	1263.4	b	–
Asp-270–Lys-282	1287.5	13 a	1286.5	11 a	1287.0
Asn-89–Arg-97	1064.1	b	1064.2	b	1064.6
Ser-262–Lys-269	946.1	14 a	947.7	12 a	946.8
Asp-270–Lys-282	1287.5	b	1288.2	b	–
Thr-256–Glu-267	1421.5	15 a	1422.3	13 a	1422.4
Leu-163–Leu-174	1379.6	b	1378.3	b	–
Asp-1–Glu-16	1752.0			14	1751.7
Gln-124–Lys-138	1724.0	16	–	15	1723.5
Ser-46–Lys-75	3679.2	17	–	16	3678.7
Ala-191–Lys-218	3521.2			17	3520.0
Ala-191–Phe-214	3123.6			18	3122.7

For the sequence-derived molecular mass calculation, the residual masses (monoisotopic values) of the individual amino acids were used (–, not found in mass spectrum).

Table 2
Mass spectrometric data for peptides resulting from the first partial acid hydrolysis by MALDI analysis after RPLC separation (A) and ESMS analysis after LCMS separation (B)

(A) Fractions from SMART

Peptide	Sequence position	Calculated mass	MALDI
2 a	Asn-151–Ala-180	3240.6	3241.6
b	Leu-101–Val-113	1340.5	1341.9
3 a	Val-289–Glu-323	3848.4	3851.6
b	Pro-41–	–	–
4 a	Asp-288–Glu-323	3963.5	3963.3
b	Ala-135–Val-152	1979.3	1980.0
c	Asn-64–	–	–
5	Val-289–Gly-301	1385.6	1386.2
6	Ala-271–Glu-323	5642.4	5641.7

(B) Fractions from LCMS

Peptide	Sequence position	Calculated mass	ESMS
1	Gly-96–Lys-99	430.5	430.8
2 a	Ala-222–Gly-228	629.7	630.6
b	Ser-168–Asp-171	464.5	465.4
3	Leu-172–Asp-178	822.9	822.6
4	Ala-135–Phe-149	1651.0	1651.2
5 a	Ala-271–Pro-286	1581.9	1582.2
b	Ala-271–Asp-287	1697.0	1698.0
6	Ala-2–Phe-39	4351.0	4351.6
7 a	Val-289–Glu-323	3848.4	3848.9
b	Ala-135–Asp-167	3538.0	3537.8

Sequence derived masses were calculated as stated in Table 1.

however, leave no doubt that the sequence given in Fig. 1 is correct. Our work is in agreement with the sequence deduced from the gene [3], including as a major point, the 19 residue insertion (Asn-79–Arg-97). At one position, residue 152, we find a valine whereas it is a methionine in the gene sequence. The presence of a Val was unambiguously proven by sequence analysis of peptides AH1 SM 2a and AH2 4d as well as by mass spectrometric analysis of peptides AH1 LCMS 7b, AH1 SM 2a and AH1 SM 4b (Table 2). It is possible that the difference may be due to the use of a different strain in both investigations. Whereas we used strain ATCC 10145 (=NCTC 10332), Ridout et al. used strain ATCC 19429 (=NTCC 6750).

Regardless of the identity of residue 152, the new sequence has important implications for the model of the mechanism of activity of PaCCP. Ellfolk et al. proposed a model based on the assumption that the polypeptide chain of PaCCP is folded as two linked single haem-binding domains of the mitochondrial cytochrome c type (Fig. 3A) [8]. Considering the standard spacing of 60 residues between the fifth and the sixth haem ligand for mitochondrial class I cytochromes, they did not see a putative sixth haem ligand for the first haem group within the N-terminal domain. On the basis of NMR data, which suggest that the two haems would be in close contact to each other, the authors proposed that His-240 (261 in Fig. 1), situated in the C-terminal domain, might act as the sixth ligand and would therefore confer a low potential character to the first haem. Taking into account the 19 residues insertion in the N-terminal

domain, we now see that Met-115, located 59 residues from His-55, becomes a very likely candidate as second haem ligand. Consequently, we think that His-261, now freed from its proposed role as ligand in the N-terminal domain, is the partner ligand for His-201 in the C-terminal domain again with a separation of 59 residues. This new model is shown in Fig. 3B. It confers a high redox potential to the N-terminal domain which interacts with the electron donor, cytochrome c-551, and a low redox potential to the C-terminal domain where the peroxide is bound. Alternatively, Met-275, instead of His-261, might act as the second ligand for the C-terminal domain. Bacterial cytochromes c_2 of Class I are known where the spacing between the two ligand residues is up to 80 residues (as e.g. in *Paracoccus denitrificans* cytochrome c-550 [9]). However, since a methionine as the sixth ligand is nearly always associated with a high redox potential of the haem, this interpretation would imply that PaCCP has two haems with high redox potential which is in contrast with the observation that the haems act at redox potentials differing as much as 600 mV [2]. Consequently, this second possibility is less likely although there is a precedent for a low potential haem with a His-Met ligand structure [10]. Finally, we can not exclude the possibility of a mechanism in which Met-275 ligates the second haem site while His-116 ligates the first haem. The relative configuration of the high and low redox potential sites would then be as originally suggested by Ellfolk et al. [2].

Final proof for any of these proposals will of course have to await three-dimensional structure analysis of PaCCP but in the meantime the proposal for the mechanism of action of PaCCP is more likely to be that of Fig. 3B rather than that of Fig. 3A.

Acknowledgements: This work has been supported by the Belgian Fund for Joint Basic Research (32001891) and a Concerted Research Action (12052293) from the Flemish Government.

References

[1] Rönnberg, M., Kalkkinen, N. and Ellfolk, N. (1989) FEBS Lett. 250, 175–178.
[2] Ellfolk, N., Rönnberg, M. and Österlund, K. (1991) Biochim. Biophys. Acta 1080, 68–77.
[3] Ridout, C.J., James, R. and Greenwood, C. (1995) FEBS Lett. 365, 152–154.
[4] Foote, N., Turner, R., Brittain, T. and Greenwood, C. (1992) Biochem. J. 283, 839–843.
[5] Klarskov, K., Roecklin, D., Bouchon, B., Sabatié, J., Van Dorsselaer, A. and Bisschoff, R. (1994) Analyt. Biochem. 216, 127–134.
[6] Hillenkamp, F., Karas, M., Beavis, R.C. and Chait, B.T. (1991) Anal. Chem. 63, 1193A-1203A.
[7] Samyn, B., Berks, B.C., Page, M.D., Ferguson, S.S. and Van Beeumen, J.J. (1994) Eur. J. Biochem. 219, 585–594.
[8] Moore, G.R. and Pettigrew, G.W. (1990) Cytochromes C: Evolutionary, Structural and Physicochemical Aspects, Springer, Berlin, Germany.
[9] Ambler, R.P., Meyer, T.E., Kamen, M.D., Schichman, S.A. and Sawyer, L. (1981) J. Mol. Biol. 147, 351–356.
[10] Mc Lendon, G.L., Bagby, S., Charman, J.A., Driscoll, P.C., McIntire, W.S., Mathews, F.S. and Hill, H.A.O. (1991) Proc. Natl. Acad. Sci. USA 88, 9463–9467.