Biochimica et Biophysica Acta, 1058 (1991) 56–60

ADONIS 000527289100129V

BBABIO 43405

Primary structure diversity of prokaryotic diheme cytochromes *c*

J. Van Beeumen

Laboratory of Microbiology and Microbial Genetics, State University of Ghent, Ghent (Belgium)

(Received 5 February 1991)

Key words: Primary structure; Sequence; Cytochrome *c*; Cytochrome, diheme; Diheme cytochrome

The primary structure of five diheme cytochromes from photosynthetic bacteria recently determined in our laboratory lead to the first insights in the structural diversity of this type of cytochrome. A schematic overview is given, relating these structures to the four diheme cytochrome sequences already available. The comparison reveals unexpected homologies.

In contrast to the rather uniform structure of eukaryotic mitochondrial cytochromes, the prokaryotic cytochromes c are very diverse in structure. Sequence analyses over the past 20 years, many of them performed by Ambler [1], have led to a classification of prokaryotic cytochromes into three main classes based in part on the number and the location of the heme groups. Class I and class II cytochromes have a single heme bound respectively near the N- and the C-terminal end of a roughly 100 (± 20) residues long polypeptide, whereas class III cytochromes are characterized by having 3 to 4 hemes bound to a chain of 68–111 residues. The latter class of proteins appear to occur only in sulfate and in some sulfur-reducing bacteria [2]. The three-dimensional structure of nine prokaryotic cytochromes have been determined, four from class I, one from class II and four from class III (referred to in Refs. 1–3). Recently, also the structure of the tetraheme cytochrome *c* bound to the photosynthetic reaction center of *Rhodopseudomonas viridis* has been determined [4]. There is no sequence homology to the tetraheme cytochromes of the sulfate-reducing bacteria. The heme binding site of all prokaryotic cytochromes *c* is of the sequence type Cys-X-Y-Cys-His, except for two of the four sites in three cytochromes c_3 where there are four residues between the two cysteines (see references in Ref. 2).

Very few structural data are known for cytochromes having two covalently bound hemes. The reason herefore may be that many of these proteins seem to occur in rather small amounts. The only complete diheme cytochrome *c* sequences published are those of *Azotobacter vinelandii* cytochrome c_4 [5] and of *Pseudomonas aeruginosa* cytochrome *c* peroxidase [6]. Also, the nearly complete sequences of *P. aeruginosa* cytochrome c_4 [5] and of *P. stutzeri* cytochrome *c*-552 [7] are known. To this short list I can now add the complete sequences of four cytochromes, and the nearly completed sequence of a fifth, all isolated from phototrophic bacteria. The general outline of these structures is given in Fig. 1. In order to discuss and relate them to the structures of the published diheme cytochromes I shall first focus briefly on the main properties of the latter.

Cytochrome c_4 from *A. vinelandii* is 190 amino acids long and has the first of the two heme groups bound near the N-terminus of the polypeptide chain. Both are heme-binding sites of the classical Cys-X-Y-Cys-His type. There are 101 residues between the second cysteine of the first and the first cysteine of the second site. Upon aligning the two heme binding sites a sequence similarity of 24% between the two halves of the cytochrome can be calculated. This figure suggests that the diheme cytochrome originated from gene duplication of a single small monoheme class I cytochrome. A typical sixth heme ligand methionine which is found in high redox potential monoheme cytochromes also occurs in both halves of the protein at homologous positions (Fig. 1). Cytochrome c_4 from *P. aeruginosa* has the same overall structural features as the protein from *Azotobacter*. There is 76% sequence homology with the latter protein, which is comparable to the 61% found between the monoheme cytochromes *c*-551 of the two organisms. Cytochrome c_4 from *P. aeruginosa* is under study by X-ray crystallography [8]. Although a detailed structure has not yet been published, preliminary data reveal that the protein is composed of two covalently bound monoheme cytochromes which each have the

Correspondence: J. Van Beeumen, Laboratory of Microbiology and Microbial Genetics, State University of Ghent, Ledeganckstraat 35, 9000 Ghent, Belgium.

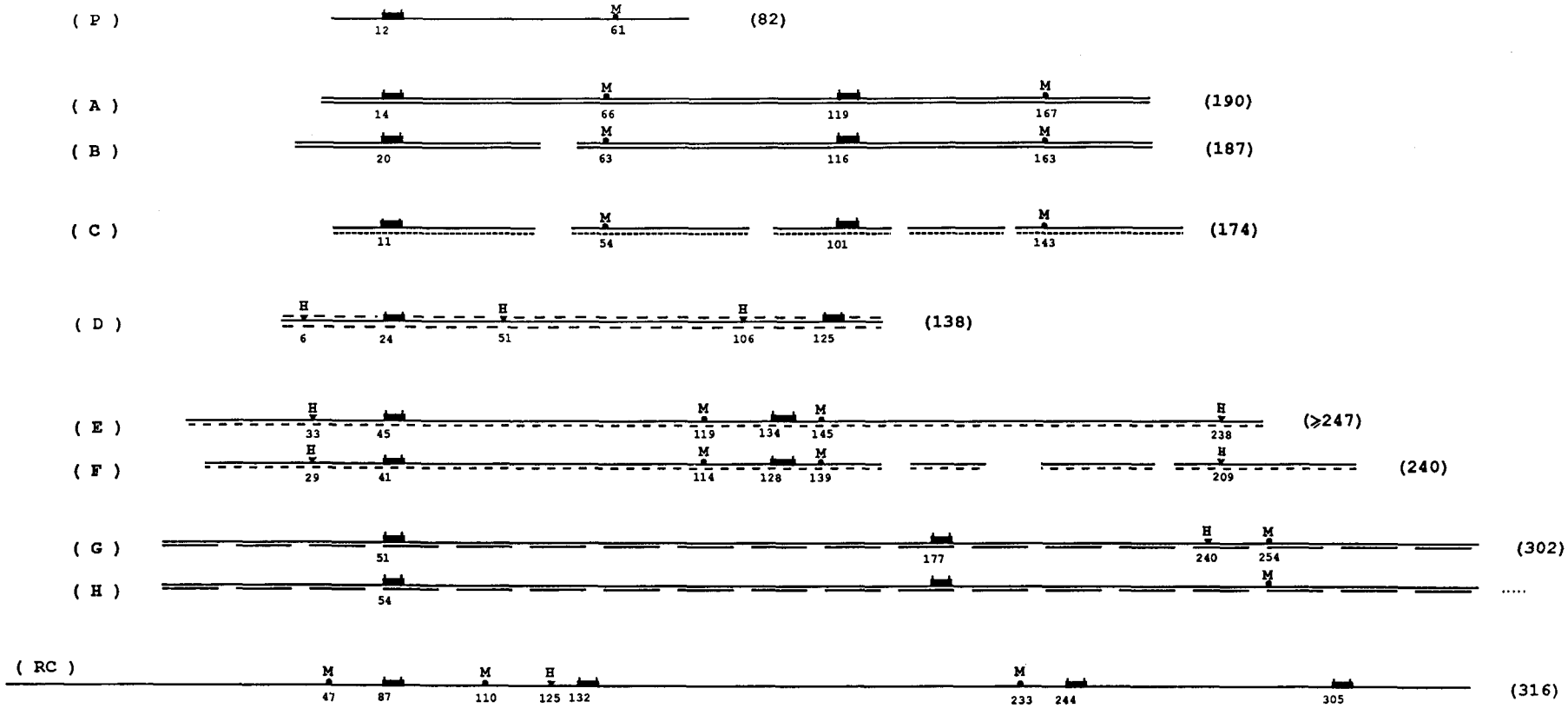

Fig. 1. Schematic presentation of diheme cytochromes *c*.

(A) *Azotobacter vinelandii* cytochrome c_4;
(B) *Chromatium vinosum* cytochrome *c*-553(550);
(C) *Chromatium vinosum* flavocytochrome *c*, heme subunit;
(D) *Rhodobacter sphaeroides* cytochrome *c*-551.5;
(E) *Pseudomonas stutzeri* cytochrome *c*-552;
(F) *Rhodocyclus tenuis* cytochrome *c*-551;
(G) *Pseudomonas aeruginosa* cytochrome *c* peroxidase;
(H) Rhodobacter capsulatus cytochrome *c*-556.

In relation to the discussion in the text we have also included the monoheme cytochrome *c*-551 from *Pseudomonas aeruginosa* (P) and the tetraheme cytochrome from the reaction center of *Rhodopseudomonas viridis* (RC). The heme binding sites are indicated as horizontal blocks; presumed sixth ligand residues are shown as filled circles (methionine) or triangles (histidine). The first hemes of the various sequences are aligned with each other. The length of each line is proportional to the number of residues in each sequence. This number is given between brackets at the end of each line. Deletions appear as open spaces. Sequences that are considered to be homologous are represented with similar type of line.

'cytochrome fold' of the class IC cytochrome *c*-551 from *P. aeruginosa* [9]. The precise function of the diheme cytochromes c_4 remains unknown at present.

The polypeptide of cytochrome *c* peroxidase from *P. aeruginosa* contains 302 residues with heterogeneity at six positions, suggesting more than one gene [6]. The two heme-binding sites are of the classical type and they are separated by 122 residues (Fig. 1). In contrast to the hemes of the cytochromes c_4, which both have a high redox potential, the heme group bound near the N-terminus of the peroxidase has a low potential, and the one in the middle of the molecule has a high redox potential. The distal ligand for the latter heme is proposed to be Met-254, whereas weak and by no way proven arguments are given to consider His-240 as the sixth ligand of the low-potential heme (at low temperature). There is no internal sequence homology in the *Pseudomonas* enzyme and it is apparently not related to any of the four classes of *c*-type cytochromes.

A third type of diheme cytochrome has been isolated from anaerobically grown cells of *P. stutzeri* (previously called *P. perfectomarina* [10]). It is a cytochrome *c*-552 very likely involved in denitrification. The primary structure of cytochrome *c*-552 [7] is different from the cytochromes c_4 in several ways. Firstly, the N-terminal heme-binding site, of the classical type, is as far as residues 45 and 48. Secondly, the sequence preceding this site contains two tryptophan residues. No other bacterial class I cytochrome *c* has even a single tryptophan ahead of the first cysteine residue except for the cytochrome subunit of *P. putida* flavocytochrome [11]. Thirdly, and most surprisingly, the second heme-binding site contains three residues between the two cysteines and is thus of the type Cys-X-Y-Z-Cys-His. Moreover, X is a tryptophan residue. As yet, an aromatic residue between the cysteines has never been found in any *c*-type cytochrome except in the reaction center cytochrome *c* from *Rps. viridis*, which contains a Phe or a Tyr in three out of the four heme binding sequences [12]. Finally, there are 85 amino acids between the two heme-binding sites in the cytochrome *c*-552, 17 residues less than in the cytochrome c_4. All these specific sequence characteristics suggest that the overall three-dimensional structure of this protein is likely to be quite different from that of the *Pseudomonas* type IC cytochromes. The two hemes of cytochrome *c*-552 have measurably different redox potentials, of +174 and −180 mV [13]. By analogy with the high redox potential class I cytochromes, one is inclined to believe that the N-terminal heme is the one with the positive potential. It is not obvious from the sequence, however, which residue would then be the sixth ligand. There is a methionine 70 residues down the sequence of the fifth ligand His-49, but such a distance is too long for this methionine to be the sixth ligand of a IC cytochrome. The physiological role of the cytochrome *c*-552 is unknown, but it has been proposed, based on the susceptibility to proteolytic cleavage, that the protein would act as a cytochrome *c* peroxidase, thereby protecting nitrite reductase from being denatured if growth conditions were to become aerobic [7].

Cytochrome *c*-551 from *Rhodocyclus tenuis*, the first of the five diheme cytochromes we have been studying over the past 2 years, is a protein with two hemes of low redox potential (Meyer, T.E., personal communication). There are 240 residues in the polypeptide chain with one classical heme-binding site (starting at Cys-41) and the other one (starting at Cys-128) having the sequence Cys-Trp-Gly-Ser-Cys-His. As indicated above, these characteristic features have until now been found only in cytochrome *c*-552 from *P. stutzeri*. Moreover, cytochrome *c*-551 from the photosynthetic bacterium also displays the other specific sequence features of cytochrome *c*-552 such as the occurrence of two tryptophan residues in the N-terminal region and a total of 83 residues between the first and the second heme-binding site. Alignment of the cytochromes *c*-551 and *c*-552 only requires the assumption of six 1 or 2 residue deletions in the first half and of three deletions of respectively 6, 12 and 5 residues in the C-terminal region of the proteins (Fig. 1). The overall similarity between the two cytochromes is 39%, a number which will not change very much when the complete sequence of the *Pseudomonas* cytochrome becomes available. It is not obvious to point out which are the distal ligands of the two hemes in cytochrome *c*-551. The residues Met-120, Met-146, His-33 and His-239 of the cytochrome *c*-552 are invariant in the cytochrome *c*-551, yet the *Rc. tenuis* protein does appear to have only low redox potential hemes. The presence of this unusual type of diheme cytochrome in a photosynthetic and a denitrifying prokaryote may seem rather remarkable. One should keep in mind, however, that both organisms also contain a soluble cytochrome *c* of subclass IC (cytochromes *c*-553 and *c*-551, respectively) which have 53% sequence homology. Two other purple non-sulfur bacteria, *Rc. gelatinosa* and *Rc. purpureus*, also contain such a small class IC cytochrome; all other Rhodospirillaceae, as, for example, *Rb. sphaeroides* and *R. molischianum*, have a longer cytochrome of subclass IA or IB [1]. The occurrence of the same class IC cytochrome in *Rc. tenuis* and *P. stutzeri* is part of the arguments that induced the theory according to which the aerobic bacteria of the *Pseudomonas* type originated from an ancestral phototroph that evolved to the present-day bacteria of the type *Rhodocyclus* [14]. The case of the diheme cytochromes of *Rc. tenuis* and *P. stutzeri* can be considered as additional evidence for this theory.

The low redox potential cytochrome *c* of *Rb. sphaeroides* (−254 mV), a cytochrome *c*-551.5, is a water-soluble protein of 138 amino acids (Fig. 1). This is only a few residues more than the longest class IA

cytochrome (*Paracoccus denitrificans* c_2 [1]), yet the protein has two heme-binding sites. In contrast to cytochrome *c*-551 from *Rc. tenuis*, both sites are of the classical type. Even more surprising, however, is that the second site occurs near the C-terminus of the polypeptide chain starting at position 12. The cysteines of the first site occur at the positions 24 and 27. There is no obvious sequence homology to any other cytochrome *c*. The protein has an acidic isoelectric point due to the preponderance of ten acidic residues. Secondary structure predictions and circular dichroism measurements suggest the presence of six short α-helices for a total amount of 23%, a feature unknown to any other cytochrome *c*. Based on these data, we are forced to conclude that cytochrome *c*-551.5 from *Rb. sphaeroides* is the first example of a new structural class of cytochromes. Possible sixth ligands are the histidines at positions 6, 51 and 106.

Cytochrome *c*-553(550) from the purple sulfur bacterium *Chromatium vinosum* displays a split-α peak in the reduced form and has a positive redox potential of 330 mV. It is a membrane-bound protein which has been solubilized with an aqueous acetone solution [15]. The polypeptide chain of this protein contains 187 residues with two heme binding sites of the classical type starting at the positions 20 and 116 (Fig. 1). Unlike the sequence of *Rb. sphaeroides* cytochrome *c*-551.5, which so far seems to be a unique structure, cytochrome *c*-553(550) is undoubtedly homologous to the cytochromes c_4 from *A. vinelandii* and *P. aeruginosa* with a similarity of 48–51%. Such a high similarity is totally unexpected, given the strongly different nature of the energy generating electron-transport chains of both bacterial species. The role of the *Chromatium* cytochrome is unknown, but the high redox potential suggests that the protein may be an electron donor to the photosynthetic reaction center. The sequence similarity between the two halves of the cytochrome is 30% which is 6% higher than for the two halves of *Azotobacter* cytochrome c_4. The similarity between corresponding halves of the two cytochromes c_4 amounts to 48 and 50% which suggests divergence in a synchronous way. Cytochrome c' from *C. vinosum* is the only other cytochrome from this species that has been sequenced [16]. It is a class II cytochrome of which members have been found in other phototrophic purple bacteria and in the aerobic bacteria *Agrobacterium tumefaciens* and one species of *Alcaligenes*, but not yet in a species of the pseudomonads.

A second diheme cytochrome from *C. vinosum* recently sequenced by our group is the cytochrome *c* subunit of the flavocytochrome *c*, a protein complex very likely acting as a sulfide dehydrogenase. With its 174 residues it is about double the size of the cytochrome *c* subunit from *Chlorobium thiosulfatophilum* flavocytochrome *c* [17]. The two heme-binding sites of the *Chromatium* subunit are both of the classical type, but there are only 86 residues between both sites and not 101 as in *A. vinelandii* cytochrome c_4. The connecting peptide between the two halves of the subunit is a hexapeptide, whereas it is 16 residues long in the cytochrome c_4. Moreover, there is only 20% sequence similarity of the first half of the *Chromatium* cytochrome with either half of the *Azotobacter* cytochrome and for the second half it is even as little as 10%. These numbers make is very unlikely that the two proteins evolved from a common ancestral structure. The first half of the *Chromatium* subunit is 38% similar to the *Chlorobium* subunit, whereas for the second half this is only 11%. The two halves of the purple sulfur cytochrome have only 10% identity. Finally, the *Chromatium* cytochrome is 6–10% less similar to the cytochrome subunit of the flavocytochrome enzyme *p*-cresol methyl hydroxylase from *P. putida* [11], than is the *Chlorobium* subunit (25%).

The last protein we are currently studying, a cytochrome *c*-556, is a diheme cytochrome from *Rb. capsulatus*. The protein has been detected in a soluble form in supernatants of the bacterium during experiments in which the electron donors to the photochemical reaction center have been probed by deleting the genes for cytochromes *b*, c_1 and/or c_2 [18]. The total length of the high redox potential cytochrome *c*-556 (+316 mV) is not completely known, but we have sequence evidence for nearly 300 residues. There is no sequence similarity to the purple bacterial cytochromes but, completely unexpectedly, there is as much as 45% average similarity with cytochrome *c* peroxidase from *P. aeruginosa*. For two regions of 20 residues this number is even 75%. The possible ligand Met-254 of the peroxidase [6] is invariant in the cytochrome *c*-556, but whether the same is true for the possible sixth ligand His-240 of the low redox potential heme of the peroxidase remains to be confirmed.

In summary, the schematic data on the five proteins discussed briefly above indicate, at least for the cytochromes *c*-553(550) from *C. vinosum*, *c*-551 from *Rc. tenuis*, and *c*-556 from *Rb. capsulatus*, that they are homologous to cytochromes present in denitrifying pseudomonads. Although the phylogenetic importance of this discovery must await sequence information for cytochrome species other than those described, it does appear from our data that 'gene transfer', proposed to be an explanation for the erratic distribution of different cytochromes in very different organisms, may be an evolutionary event that is even more frequent than previously anticipated [1]. Alternatively, the data can also be explained as if there has been no gene transfer at all. The phylogenetic relatedness between *Chromatium* and *Pseudomonas* as deduced from the cytochromes c_4 has also been detected by comparing 16S rRNA sequences [19].

A central question in relation to the multiheme cytochromes in general and to the diheme cytochromes in particular is whether they originated from gene duplication or, alternatively, from a fusion of small cytochrome genes. Although we do not have enough data yet to propose a certain lower limit of similarity for gene duplication as the evident explanation, we can assume that such low values as the 10% for the two halves of the *Chromatium* flavocytochrome subunit are a good indication that the protein must be the result of gene fusion. The cytochromes c_4 from *C. vinosum* and *P. aeruginosa*, on the other hand, are clearly examples of a process of gene duplication. For the cytochromes *c*-551.5 from *Rb. sphaeroides* and the cytochromes *c*-552 from *P. stutzeri* and *Rc. tenuis* it is more difficult to reach a conclusion: for the first case because of the location of a binding site near each end of the polypeptide chain, for the second because of the different structure of the two binding sites and the absence of any similarity overall. The sequence evidence of the cytochromes c_3 from *Desulfovibrio* and the reaction center cytochrome from *Rm. viridis* suggests that they have originated from two successive gene doublings (Fig. 1). The new sequences that we have schematically described or the three that were known before provide evidence, however, that they could be a precursor of the tetraheme cytochromes.

Acknowledgement

This work is supported by the Belgian national incentive program on fundamental research in life sciences initiated by the Belgian State – Prime Minister's Office – Science Policy Programming Department (Contract BIO 22).

References

1 Ambler, R.P. (1991) Biochim. Biophys. Acta 1058, 42–47.
2 Moura, J.J.G., Costa, C. Liu, M.J., Moura, I. and LeGall, J. (1991) Biochim. Biophys. Acta 1058, 61–66.
3 Dickerson, R.E. (1980) Sci. Am. 242, 98–110.
4 Deisenhofer, J., Epp, O., Miki, K., Huber, R. and Michel, H. (1985) Nature 318, 618–624.
5 Ambler, R.P., Daniel, M., Melis, K. and Stout, C.D. (1984) Biochem. J. 222, 217–227.
6 Ronnberg, M., Kalkkinen, N. and Ellfolk, N. (1989) FEBS Lett. 250, 175–178.
7 Denariaz, C.M., Liu, M.Y., Payne, W.J., LeGall, J., Marquez, L., Dunford, H.B. and Van Beeumen, J. (1989) Arch. Biochim. Biophys. 270, 114–125.
8 Sawyer, L., Jones, C.L., Damas, A.M., Harding, M.M., Gold, R.D. and Ambler, R.P. (1981) J. Mol. Biol. 153, 831–835.
9 Almassy, R.Y. and Dickerson, R.E. (1978) Proc. Natl. Acad. Sci. USA. 75, 2674–2678.
10 Liu. M.C., Payne, W.J., Peck, H.D. and LeGall, J. (1983) J. Bacteriol. 154, 278–286.
11 McIntire, W., Singer, T.P., Smith, A.J. and Mathews, F. (1986) Biochemistry 25, 5975–5981.
12 Weyer, K.A., Lottspeich, F., Gruenberg, H., Lang, F., Oesterhelt, D. and Michel, H. (1987) EMBO J. 6, 2197–2202.
13 Liu, M.C., Peck, H.D., Payne, W., Anderson, J.L. Dervartanian, D.V. and LeGall, J. (1981) FEBS Lett. 129, 155–160.
14 Dickerson, R.E. (1980) Nature 283, 210–212.
15 Cusanovich, M.A. and Bartsch, R.G. (1969) Biochim. Biophys. Acta 189, 245–255.
16 Ambler, R.P., Meyer, T.E. and Kamen, M.D. (1979) Nature 278, 661–662.
17 Van Beeumen, J., Van Bun, S., Meyer, T.E., Bartsch, R.G. and Cusanovich, M.A. (1990) J. Biol. Chem. 265, 9793–9799.
18 Prince, R.C. and Daldal, F. (1987) Biochim. Biophys. Acta 894, 370–378.
19 Woese, C.R. (1987) Microb. Rev. 51, 221–271.