ELSEVIER

Biochimica et Biophysica Acta 1231 (1995) 220–222

BBA Biochimica et Biophysica Acta

Short Sequence-Paper

Molecular cloning and sequencing of cytochrome *c′* from the phototrophic purple sulfur bacterium *Chromatium vinosum*

Mark T. Even [a], Richard J. Kassner [a,*], Marti Dolata [b], Terrance E. Meyer [b], Michael A. Cusanovich [b]

[a] *Department of Chemistry, University of Illinois at Chicago, Chicago, IL 60607, USA*
[b] *Department of Biochemistry, University of Arizona, Tucson, AZ 85721, USA*

Received 2 May 1995; accepted 20 June 1995

Abstract

The gene for cytochrome *c′* from *Chromatium vinosum* was cloned from a *Hin*dIII-*Sal*I digest of genomic DNA. A 1.4 kbp fragment containing the gene was sequenced in both directions using the Sanger dideoxy method. The cytochrome *c′* gene codes for a 154-residue peptide, of which the last 131 amino acids match the previously determined sequence of the protein. The remaining 23 residues represent a signal sequence that is cleaved from the polypeptide upon translocation to the periplasmic space. An additional open reading frame on the other strand of the fragment codes for a peptide that contains four regions that are homologous to corresponding regions of the cytochrome *b*-type subunit of several Ni-Fe hydrogenases.

Keywords: Cytochrome *c′*; Nucleotide sequence; (*C. vinosum*)

Cytochromes *c′* are a class of heme proteins isolated from a wide variety of denitrifying and photosynthetic bacteria [1]. They are usually isolated as dimers of identical 14 kDa subunits, each containing a covalently bound heme at a Cys-X-X-Cys-His site near the carboxy terminus [2]. High-resolution crystal structures have been published for cytochromes *c′* from *Rhodospirillum molischianum* [3], *Rhodospirillum rubrum* [4], and most recently for *Chromatium vinosum* [5], with all three having similar structures. Each subunit is a left-twisted, four-α-helical bundle, with the heme in the center, two helices on either face, with a histidine as the fifth ligand to the iron, and the open sixth iron-coordination site buried and facing the subunit interface. The physical-chemical properties of the cytochromes *c′* are unique among the high-spin hemoproteins. The spin-state of the ferric cytochromes *c′* has been described as a quantum mechanical admixture of intermediate and high-spin states [6], while the ferrous cytochromes *c′* are high-spin similar to hemoglobin and myoglobin [7]. The cytochromes *c′* exhibit only limited binding of some common heme ligands, including CO, NO, CN^- and various alkyl-isocyanides [8]. *C. vinosum* cytochrome *c′* reversibly dissociates from dimer to monomer upon ligand binding, a behavior not observed by the cytochromes *c′* from most other organisms [9,10]. Amino-acid sequences for many cytochromes *c′* have been determined [11], but the nucleotide sequence reported here is the first for the cytochromes *c′*.

C. vinosum strain D was grown on a malate containing medium [12], and the genomic DNA isolated by the CTAB method [13]. A subgenomic library was constructed in pBluescriptIIKS + (Stratagene) from *Hin*dIII-*Sal*I digested *C. vinosum* genomic DNA. This library was screened with mixed oligonucleotide probes corresponding to two different regions of the 131-residue mature protein. A positive double-stranded clone containing a 1.4 kbp insert was isolated and sequenced in both directions using the dideoxy terminator method [14] with US Biochemical's Sequenase and Taquence Version 2.0 kits. Due to numerous areas of secondary-structure, the insert was subcloned into pBluescriptIISK + . Single-stranded DNA was isolated and sequenced from single-stranded rescues from both the subclone and the original clone phagemids [15].

The determined nucleotide sequence (Fig. 1) contains two open reading frames. The cytochrome *c′* gene encodes a 154-residue protein that matches the previously determined sequence of the 131-amino-acid mature protein [16],

* Corresponding author. Fax: + 1 312 9960431.

SSDI 0005-2728(95)00101-8

```
                                                    Orf 1
1     AAGCTTTTCCATCCACATCCATCGACCAACAACGGAGATTGTCCGATGAAACACGTGTTAGCCAGCACCGCCGCCGGTCTCATGGCCCTT
1                                                   MetLysHisValLeuAlaSerThrAlaAlaGlyLeuMetAlaLeu

91    GGCCTCGCCAGCTCTGCCATCGCCGCCGGCCTCTCGCCCGAGGAGCAGATCGAGACCCGTCAGGCCGGCTACGAATTCATGGGCTGGAAC
16    GlyLeuAlaSerSerAlaIleAlaAlaGlyLeuSerProGluGluGlnIleGluThrArgGlnAlaGlyTyrGluPheMetGlyTrpAsn
                          ⇑
181   ATGGGCAAGATCAAGGCGAACCTGGAGGGCGAATACAACGCCGCTCAGGTCGAGGCCGCCGCCAACGTCATCGCCGCCATCGCCAACTCC
46    MetGlyLysIleLysAlaAsnLeuGluGlyGluTyrAsnAlaAlaGlnValGluAlaAlaAlaAsnValIleAlaAlaIleAlaAsnSer

271   GGCATGGGCGCGCTCTATGGTCCGGGCACCGACAAGAATGTCGGCGACGTGAAGACCCGCGTGAAGCCCGAGTTCTTCCAGAACATGGAA
76    GlyMetGlyAlaLeuTyrGlyProGlyThrAspLysAsnValGlyAspValLysThrArgValLysProGluPhePheGlnAsnMetGlu

361   GACGTCGGCAAGATTGCGCGTGAGTTCGTCGGTGCGGCCAACACGCTGGCTGAAGTGGCCGCCACCGGTGAGGCCCAGGCGGTCAAGACC
106   AspValGlyLysIleAlaArgGluPheValGlyAlaAlaAsnThrLeuAlaGluValAlaAlaThrGlyGluAlaGluAlaValLysThr

451   GCTTTCGGCGATGTCGGCGCCGCGTGCAAATCGTGCCACGAGAAGTATCGGGCCAAGTGATCCCGGGGCGATTTCCCGAAAAGAGTGGGA
136   AlaPheGlyAspValGlyAlaAlaCysLysSerCysHisGluLysTyrArgAlaLys***

541   CAAGCAGAGCGTAGCGATGCCGGCTTACTCGGTGGGGAATCGTCGTTAGAGATCCGCACCCGATCATGCAGAAGGCCGCCCAAGAAGAAT

                          ***TrpAspProValSerGlyThrProAlaProProProProSerLeuLeuGlyGlyAsnAlaIleTrp
631   GCCCGCCGGAAAAACGCAAGCAGTGGTTAGGCCATGCGAAGGGCAGCCGCGCCCACCGCCTCCGCTGTCGTCAGGCGGTAACCGCTAGGT

      ValAlaGlyAlaThrIleAlaLeuAlaLeuValPheAlaLeuLeuGlyGlyGlyLysAlaSerHisAlaMetProGluArgThrMetLys
721   CTGGCGCGGCCGGCACTACCGCTCGCGGTCGTGCTTGCGCTCGTCAGGTGGTGGGAACCGGCTCACGCGGTAGCCGAGCGCCCAGTAGAA

      TyrGlyThrLeuMetProArgIleLeuAsnAspGlyArgValTrpThrHisPheLeuThrAlaLeuValHisLeuGlyValPheGlyGly
811   CATTGGCCACTCGTACCCGGCCTAGTCCAACAGTGGTGCGTGGGTCCATACCTTTTCACAGCGGTCGTGTACGTCCGGCTGCTTCGGTGG

      IleIleTrpIleAlaGlnArgHisLeuSerSerLeuTrpValSerThrGluSerSerValLeuSerLysLeuProGlySerPheAlaIle
901   CTACTAGGTCTACCGGACTGCCACGTCTGACGACTCGGTCTGCGAGCAGAGCGACCTCTGGTCACTGAACTCGCCGGGCCTCTTCCGCTA

      AspAspThrAlaPheLeuGlyThrLeuSerGlnPheLeuMetLeuAlaLeuLeuValMetValSerIleGlyGlyLeuProAsnHisGly
991   CAGCAGCCACCGCTTCTCGGGCCAGTCTGAGACCTTGTCGTAGTCGCGGTCGTCGTGGTAGTGACTCTAGGGCGGGTCGCCTAACACGGG

      MetGlyHisTrpArgGlyArgLeuTyrValLeuMetThrLysProGlyHisValPheGlnSerPheArgAlaTyrThrSerGlyIlePhe
1081  GTACGGCACGGTCGCAGGCGCGTCTATCTGCTCGTACCAGAACCCCGGCACGTGCTTGACCGACTTCGCCCGTATCCAGCTTGGCTACTT

      GlyTrpValLeuArgPheIleIleLeuGlyPheIleAlaGlnGlyIleTrpGlyHisTrpAspIleTrpArgGlyGlyThrLeuGlyThr
1171  CGGGGTCTGCTCTGCTTTCTACTAGTTAGGCTTCTACCGGACAGGCTAGGTCGGTACGGTCAGCTAGGTCGCAGGCGGGCAGTCGGGCCA

      AlaPheAlaSerValValLeuValValLeuLeuTrpHisPheLeuArgThrProLeuAspTrpLeuLysIleArgHisGluGlnMet
1261  GCGCTTCCGCCTGTGGTGGTCGTGGTGGTCGTTGGTCACTTTCTCTGCCCAGCCTTCTAGGGTCTCAAACTAAGCTACAAGGACGTAGTA
                                                                                        Orf 2
1351  AGGATATGGTCTCTCCCTACCCACCGGTTTCACAGCGGGTTAGGCCGCCAGCTG
```

Fig. 1. Nucleotide sequence of a 1.4 kbp *Hin*dIII-*Sal*I fragment of *C. vinosum* DNA. The deduced amino-acid sequences of the cytochrome *c′* gene and ORF 2 are indicated using the standard three-letter code. The coding strands are shown in the 5′ → 3′ direction from position 1 to 630 and in the 3′ → 5′ direction from position 631 to 1403. Putative ribosomal binding sites are underlined. The signal peptide cleavage site is marked by an arrow.

with the addition of a 23-residue signal peptide which is cleaved off after directing the protein to the periplasm. The sequence of this signal peptide shows similarity to the sequences of signal peptides of other periplasmic proteins from *C. vinosum* and other photosynthetic bacteria, as shown in Fig. 2. Cytochrome *c′* from *Paracoccus denitrif-*

```
1)                  M K H V L A S T A A G L M A L G L A S S A I A
2)              M T Q S T P R L M L A A S V L A L G L A S N A G A
3)    M T L N R R D F I K T S G A A V A A V G I L G F P H L A F G
4)        M D P I R T S R L R A P L L A S A L L L A G S S V C A A
5)                      M K I S I Y A T L A A I T L A L P A A A
6)                      M K K L V F G L F V L A A S V A P A A A
7)                  M K K G F L A A G V F A A V A F A S G A A L A
8)                    M K I S L T A A T V A A L V L A A P A F A
9)                    M K F Q V K A L A A I A A F A A L P A L A
```

Fig. 2. Comparison of the signal sequence of cytochrome *c′* with the leader sequences of other proteins from *C. vinosum* and other photosynthetic and denitrifying bacteria [24]. *C. vinosum*: 1, cytochrome *c′*; 2, flavocytochrome heme subunit; 3, flavocytochrome flavoprotein subunit; 4, ankyrin homolog. Cytochromes c_2: 5, *Paracoccus denitrificans*; 6, *Rhodopseudomonas viridis*; 7, *Rhodospirillum rubrum*; 8, *Rhodobacter capsulatus*; 9, *Rhodobacter sphaeroides*.

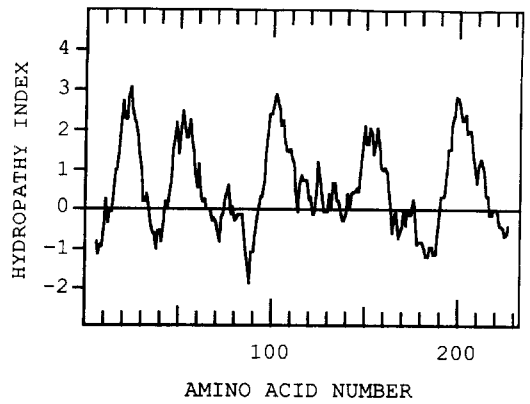

Fig. 3. Kyte and Doolittle [25] hydropathy plot of orf 2. The hydropathy plot was generated with DNA Strider.

icans was reported to be periplasmic [17], as are those from *Rhodobacter sphaeroides* and *Rhodobacter capsulatus* [18]. A low-spin homolog of cytochrome *c'*, known as cytochrome *c*-555, was cloned from *Bradyrhizobium japonicum* [19]. The *c*-555 gene also contained a leader sequence indicating that it is periplasmic. Thus, all cytochromes *c'* and homologs appear to be periplasmic.

A second open reading frame, orf 2, located on the strand opposite that of the cytochrome *c'* gene, encodes a 231-residue peptide. The hydrophobicity plot shown in Fig. 3 suggests that orf 2 may be a membrane-bound protein containing five membrane spanning regions. A BLAST search [20] of several protein and nucleotide databases indicates that orf 2 is homologous to the cytochrome *b*-type subunit of several Ni-Fe hydrogenases [21,22], with BLAST scores in the range of 32 to 77 for the membrane spanning regions, and no significant homology in the rest of the sequence. The first four membrane spanning regions of orf 2 correspond to the four membrane spanning regions in the hydrogenase subunits, including five conserved histidine residues representing potential heme ligands. However, the homology is much lower than expected based on the several known sequences of these cytochromes *b*. In fact, the gene sequence of the hydrogenase operon from the very closely related *Thiocapsa roseopersicina* indicates strong conservation of all three subunits [23]. The cytochrome *b* gene is generally followed by other genes involved in hydrogen metabolism and these do not include cytochrome *c'* [22]. Thus, the *C. vinosum* orf 2 is probably not one of the hydrogenase subunits, but could be a cytochrome *b* involved in some other pathway.

This work was supported in part by a grant from the NIH, GM21277, and the Campus Research Board of UIC. The authors wish to thank Albert S. Benight for the preparation of oligonucleotides used in the DNA sequencing.

References

[1] Meyer, T.E. and Kamen, M.D. (1982) Adv. Protein Chem. 35, 105–212.
[2] Bartsch, R.G. (1978) in The Photosynthetic Bacteria (Clayton and Sistrom, eds.), pp. 249–280, Plenum, New York.
[3] Finzel, B.C., Weber, P.C., Hardman, K.D. and Salemme, F.R. (1985) J. Mol. Biol. 186, 627–643.
[4] Yasui, M., Harada, S., Kai, Y., Kasai, N., Kusunoki, M. and Matsuura, Y. (1992) J. Biochem. 111, 317–324.
[5] Ren, Z., Meyer, T. and McRee, D.E. (1993) J. Mol. Biol. 234, 433–445.
[6] Emptage, M.H., Xavier, A.V., Wood, J.M., Alsaadi, B.M., Moore, G.R., Pitt, R.C., Williams, R.J.P., Ambler, R.P. and Bartsch, R.G. (1981) Biochemistry 20, 58–64.
[7] Maltempo, M.M., Moss, T.H. and Spartalian, K. (1980) J. Chem. Phys. 73, 2100–2106.
[8] Kassner, R.J. (1991) Biochim. Biophys. Acta 1058, 8–12.
[9] Cusanovich, M.A. and Gibson, Q.H. (1973) J. Biol. Chem. 248, 822–834.
[10] Doyle, M.L., Gill, S.J. and Cusanovich, M.A. (1986) Biochemistry 25, 2509–2516.
[11] Ambler, R.P., Bartsch, R.G., Daniel, M., Kamen, M.D., McLellan, L., Meyer, T.E. and Van Beeumen, J. (1981) Proc. Natl. Acad. Sci. USA 78, 6854–6857.
[12] Bose, S.K. (1963) in Bacterial Photosynthesis (Gest, San Pietro and Vernon, eds.), pp. 501–510, Antioch Press, Yellow Springs.
[13] Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A. and Struhl, K. (1990) Curr. Protocols Mol. Biol. 1, 2.4.1–2.4.5.
[14] Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5467.
[15] Dente, L., Sollazzo, M., Baldari, C., Cesareni, G. and Cortese, R. (1985) in DNA Cloning: A Practical Approach (Glover, ed.), pp. 101–108, IRL Press, Oxford.
[16] Ambler, R.P., Daniel, M., Meyer, T.E., Bartsch, R.G. and Kamen, M.D. (1979) Biochem. J. 177, 819–823.
[17] Gilmour, R., Goodhew, C.F. and Pettigrew, G.W. (1991) Biochim. Biophys. Acta 1059, 233–238.
[18] Bartsch, R.G., Ambler, R.P., Meyer, T.E. and Cusanovich, M.A. (1989) Arch. Biochem. Biophys. 271, 433–440.
[19] Tully, R.E., Sadowsky, M.J. and Keister, D.L. (1991) J. Bacteriol. 173, 7887–7895.
[20] Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) J. Mol. Biol. 215, 403–410.
[21] Fu, C. and Maier, R.J. (1994) Gene 145, 91–96.
[22] Vignais, P.M. and Toussaint, B. (1994) Arch. Microbiol. 161, 1–10.
[23] Colbeau, A., Kovacs, K.L., Chabert, J. and Vignais, P.M. (1994) Gene 140, 25–31.
[24] Dolata, M.M., Van Beeumen, J.J., Ambler, R.P., Meyer, T.E. and Cusanovich, M.A. (1993) J. Biol. Chem. 268, 14426–14431.
[25] Kyte, J. and Doolittle, R.F. (1982) J. Mol. Biol. 157, 105–132.