Biochem. J. (1992) 285, 255–262 (Printed in Great Britain)

Cloning, sequencing and expression in *Escherichia coli* of the rubredoxin gene from *Clostridium pasteurianum*

Isabelle MATHIEU, Jacques MEYER and Jean-Marc MOULIS
DBMS-Métalloprotéines, CENG 85X, 38041 Grenoble, France

A 3.9 kb *Bgl*II–*Hin*dIII DNA fragment containing the rubredoxin gene from *Clostridium pasteurianum* has been cloned using oligonucleotide probes designed from the protein sequence. The 2675 bp *Ssp*I–*Hin*dIII portion of this fragment has been sequenced and found to contain three open reading frames in addition to the rubredoxin gene. The putative product of one of these open reading frames is similar to various thioredoxin reductases. The rubredoxin gene translates into a sequence that differs from the previously published protein sequence in three positions, D-14, D-22 and E-48 being replaced by the corresponding amides. These changes have been confirmed by partial resequencing of the protein. Promoter-like sequences and a transcription termination signal have been found near the sequence of the rubredoxin gene, which may therefore constitute an independent transcriptional unit. Expression of *C. pasteurianum* rubredoxin in *Escherichia coli* strain JM109 has been optimized by subcloning a 476 bp *Ssp*I–*Ssp*I fragment encompassing the rubredoxin gene. Under these conditions, the latter gene was partly under the control of the *lac* promoter of pUC18, and the level of rubredoxin production could be increased twofold on addition of a lactose analogue, thus reaching 2–3 mg of pure protein/l of culture. Recombinant rubredoxin was produced in *E. coli* cells as the holoprotein, and displayed a u.v.–visible-absorption spectrum identical with that of the rubredoxin purified from *C. pasteurianum*. M.s. and *N*-terminal sequencing showed that *C. pasteurianum* rubredoxin expressed in *E. coli* differs from its native counterpart by having an unblocked *N*-terminal methionine.

INTRODUCTION

Rubredoxins are non-haem iron proteins the active site of which consists of an iron atom tetrahedrally co-ordinated to four cysteine sulphur atoms. Most rubredoxins occur in a variety of anaerobic bacteria and have a molecular mass of approx. 6 kDa. The aerobic bacterium *Pseudomonas oleovorans* produces a larger (15 kDa) rubredoxin containing two iron atoms. All of these iron proteins are presumed to serve as electron carriers, but the electron-transfer chain in which they participate has only been identified in *P. oleovorans*, where genetic analysis of the *alk* regulon has shown that rubredoxin is involved in alkane hydroxylation (Eggink *et al.*, 1987, 1990; Kok *et al.*, 1989). Several attempts have been made to define the role of rubredoxin in anaerobic bacteria. NADH–rubredoxin oxidoreductases have been at least partially purified from *Clostridium sticklandii* (Stadtman, 1965), *Desulfovibrio gigas* (LeGall, 1968) and *Clostridium acetobutylicum* (Ballongue *et al.*, 1986). Rubredoxin has also been reported to be an effective electron donor to a nitrate-reducing activity reconstituted *in vitro* from *Clostridium perfringens* (Seki *et al.*, 1988). A different approach to the identification of possible redox partners of rubredoxin, involving gene structure analysis, has been implemented with *Desulfovibrio vulgaris* (Voordouw, 1988). In the latter case, the rubredoxin gene belongs to the same operon as another gene which has been suggested to encode a rubredoxin oxidoreductase (Brumlik & Voordouw, 1989). A protein very similar to this putative gene product has been purified from *D. vulgaris* and *Desulfovibrio desulfuricans*, shown to contain two different types of iron and named desulphoferrodoxin (Moura *et al.*, 1990). Similar data are wanted for clostridia, where rubredoxin may assume a different function from that in sulphate-reducing bacteria.

In sharp contrast with the scanty knowledge of their function, rubredoxins have been structurally characterized in great detail: more than ten amino acid sequences have been reported (see alignments in Meyer *et al.*, 1990), four tridimensional structures have been elucidated with resolutions better than 0.15 nm (Watenpaugh *et al.*, 1979; Frey *et al.*, 1987; Stenkamp *et al.*, 1990; Adman *et al.*, 1991) and a considerable amount of spectroscopic data have been collected on the $Fe(SCys)_4$ active site (reviewed in Gebhard *et al.*, 1990). A strong structural basis is thus available for understanding the electron-transfer mechanisms of these proteins. Further progress will greatly benefit from the production of molecules altered by site-directed mutagenesis.

As our initial efforts towards these goals, we report here the cloning and sequencing of the rubredoxin gene from *Clostridium pasteurianum*, and the expression of the protein in *Escherichia coli*.

MATERIALS AND METHODS

Materials and strains

C. pasteurianum W5 (ATCC 6013) was obtained from the American Type Culture Collection. Competent *E. coli* DH5α cells, T_4 DNA ligase and T_4 polynucleotide kinase were from Gibco BR, Paisley, U.K. Other enzymes and *E. coli* strain JM109 were purchased from Boehringer–Mannheim. pUC18 plasmid was from Pharmacia. *E. coli* alkaline phosphatase, [γ-^{32}P]ATP and [α-[^{35}S-*thio*]]dATP were obtained from Amersham. Oligonucleotides were synthesized on a 381A Applied Biosystems machine, and those to be used as probes were end-labelled with either [γ-^{32}P]ATP or Digoxigenin-11-dUTP (Boehringer–Mannheim).

Abbreviations used: IPTG, isopropyl-1-thio-β-D-galactopyranoside; LB, Luria broth; ORF, open reading frame.

The nucleotide sequence reported in this paper has been submitted to the GenBank Nucleotide Sequence Database under the accession number M60116.

Cloning and sequencing of DNA

C. pasteurianum cells were grown on N_2 (Rabinowitz, 1972) and genomic DNA was purified as described (Saito & Miura, 1963) with a slight modification (Graves *et al.*, 1985). Genomic DNA was digested with various restriction enzymes (see the Results section), electrophoresed through agarose gels in 0.04 M-Tris/acetate/2 mM-EDTA (pH 8.5) and stained with ethidium bromide (0.3 μg/ml). DNA transfer from agarose gels to Hybond-N membranes (Amersham) was carried out with Vacugene (LKB) equipment. Prehybridization was for a minimum of 1 h at 46 °C for probe P1 and 42 °C for probe P2 (see below) in 0.1% SDS/2 × Denhardt's solution/0.75 M-NaCl/0.075 M-sodium citrate/0.1 mg of sonicated and boiled herring sperm DNA/ml. Hybridization was overnight at the same temperature in the same solution containing labelled oligonucleotide probe, and was followed by three 20 min washes at the same temperature in 0.3 M-NaCl/0.03 M sodium citrate. For cloning purposes, genomic *C. pasteurianum* DNA was digested with restriction enzymes chosen on the basis of the results of the previous hybridizations, and, without further fractionation, inserted (3:1 ratio of insert to vector) into pUC18 cut with the same or with compatible enzymes. When digestion with a single restriction enzyme was performed, the vector was also treated with *E. coli* alkaline phosphatase. Doubly digested plasmids were electrophoresed on agarose gels in order to eliminate the excised fragment of the multiple cloning site. The recombinant plasmid was used to transform *E. coli* DH5α or JM109 cells, and the cells were spread on Luria broth (LB) plates supplemented with ampicillin (50 μg/ml), 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (20 μg/ml) and isopropyl-1-thio-β-D-galactopyranoside (IPTG; 24 μg/ml). Transformants were transferred to Hybond-N membranes, either by replica plating or individually with toothpicks. Screening was carried out by colony hybridization as described above for genomic DNA hybridization. Plasmid DNA was isolated by alkaline lysis (Kieser, 1984) and sequenced by the dideoxy method (Sanger *et al.*, 1977) with the Sequenase kit (Version 2.0, United States Biochemical), and with synthetic heptadecanucleotides as primers. DNA and protein sequences were analysed using the DNASTAR software package.

Rubredoxin purification

Rubredoxin was extracted from *C. pasteurianum* cells and separated from the 2[4Fe-4S] ferredoxin on a DE-52 (Whatman) column as described previously (Moulis & Meyer, 1982). It was further purified by gel filtration on Ultrogel AcA 202 (LKB), and, if necessary, by h.p.l.c. anion-exchange chromatography on a Brownlee (Applied Biosystems) AX300 column with a linear 0.1–0.6 M-NaCl gradient in 20 mM-Tris/HCl, pH 7.4. The pure protein had an A_{490}/A_{280} ratio of 0.43 and was assayed by using a molar absorption coefficient of 8.85 $mM^{-1} \cdot cm^{-1}$ at 490 nm (Lovenberg & Sobel, 1965). Preparation of aporubredoxin, carboxymethylation of the cysteine residues, purification of peptides and sequence analysis were as described previously for *Clostridium thermosaccharolyticum* rubredoxin (Meyer *et al.*, 1990).

Rubredoxin expression in *E. coli*

Recombinant rubredoxin expressed in *E. coli* was purified as follows. One-litre cultures in LB medium were grown at 37 °C to the stationary phase (see Fig. 5 for conditions of induction with IPTG) and centrifuged (7000 *g*, 20 min). The cells were resuspended in 50 ml of 50 mM-Tris/HCl (pH 7.4)/1 mM-phenylmethanesulphonyl fluoride, disrupted by sonication, heated to 55 °C for 10 min and centrifuged (20000 *g*, 30 min). The clear pinkish supernatant was loaded on a DE52 (Whatman) column (10 cm × 2 cm) equilibrated with 0.1 M-NaCl/20 mM-Tris/HCl, pH 7.4. At this stage rubredoxin was clearly visible as a dark-red band adsorbed on top of the column. Washing with 30 ml of 0.15 M-NaCl in the same buffer eliminated yellow contaminants. Rubredoxin was eluted with 0.35 M-NaCl, concentrated in an Amicon cell fitted with a YM5 membrane, and chromatographed on a Sephadex G-50 (Pharmacia) column (100 cm × 2 cm) equilibrated with 0.1 M-NaCl/20 mM-Tris/HCl, pH 7.4. Rubredoxin fractions with $A_{490}/A_{280} > 0.30$ were pooled and further purified by h.p.l.c. on an anion-exchange column (PL-SAX; Polymer Laboratories, Church Stretton, Salop, U.K.) equilibrated with 10 mM-Tris/HCl, pH 8.0. Development with increasing salt concentrations in the same buffer eluted rubredoxin at a NaCl concentration of 0.5 M. At this stage, 2–3 mg (see the Results section) of pure rubredoxin ($A_{490}/A_{280} = 0.43$) was obtained per litre of culture.

RESULTS

Cloning of the rubredoxin gene

Analysis of the published amino acid sequence (McCarthy, 1972; Watenpaugh *et al.*, 1973) of *C. pasteurianum* rubredoxin revealed that the *N*-terminus MKKYTCT and the DDWVCP fragment were best suited for the derivation of low-degeneracy oligonucleotide probes. In addition, searches of the PIR-NBRF protein sequence database showed that these fragments were among the least likely to occur in known proteins other than rubredoxins. Taking into account the codon usage of *C. pasteurianum*, which is biased towards A or T in the wobble position (Graves *et al.*, 1985; Chen *et al.*, 1986; Hinton *et al.*, 1987; Wang *et al.*, 1988, 1990; Meyer & Gagnon, 1991), a 32-fold degenerate eicosanucleotide, 5′-GT(AG)CA(AT)GT(AG)TA-(TC)TT(TC)TTCAT-3′ (P1), and a 16-fold degenerate heptadecanucleotide, 5′-GG(AG)CA(AT)ACCCA(AG)TC(AG)TC-3′ (P2), were deduced from the MKKYTCT and DDWVCP fragments respectively. *C. pasteurianum* genomic DNA was digested with a number of restriction enzymes and analysed by Southern hybridization with P1 and P2. In any given restriction reaction, both probes hybridized with fragments of the same size. These sizes were: *Hin*dIII (4 kb), *Kpn*I (4.4 kb), *Hin*dIII–*Sac*I (2.1 kb), *Hin*dIII–*Bgl*II (3.9 kb), *Hin*dIII–*Sma*I (2.3 kb), *Hin*dIII–*Sph*I (2.3 kb). In no case was more than one fragment found to hybridize with the probes, which strongly suggests that only one rubredoxin gene is present in the genome of *C. pasteurianum*. Total genomic DNA from *C. pasteurianum* was completely digested with *Hin*dIII and *Bgl*II, and ligated into the *Hin*dIII–*Bam*HI sites of pUC18. *E. coli* DH5α cells were transformed with the recombinant DNA, plated and screened by colony hybridization with probe P1. Some 2000 transformants yielded two positive clones carrying 3.9 kb inserts which both hybridized with P1 and P2. Restriction analysis of these two clones showed them to be identical. One of the clones, the plasmid of which was designated pCPRD1, has been used in the experiments described below.

Sequence analysis

The largest part of the insert of pCPRD1, namely the 2675 bp *Ssp*I–*Hin*dIII fragment, was sequenced on both strands as shown in Fig. 1. The remainder of the insert, i.e. the approx. 1200 bp *Bgl*II–*Ssp*I fragment, was nearly completely sequenced on one strand, and only partially on the other; as it contains no remarkable features, and in particular no open reading frames, the sequence determination was discontinued, and the results are not shown here. The 2675 bp *Ssp*I–*Hin*dIII fragment contains four open reading frames (ORF), including the one encoding rubredoxin, all of them having the same orientation (Figs. 1 and

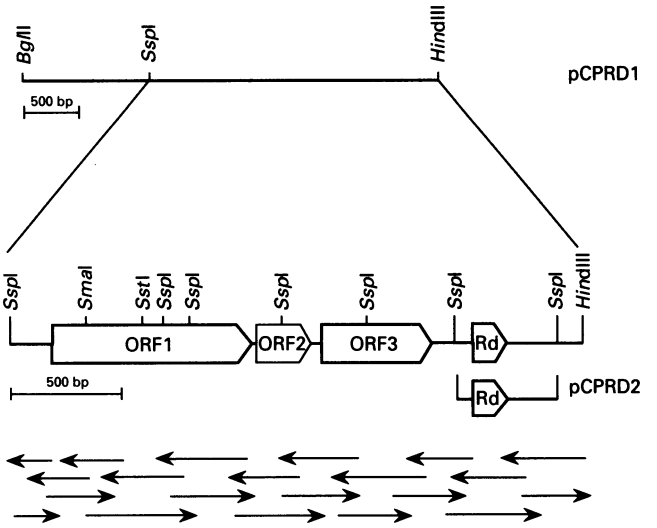

Fig. 1. Restriction map and sequencing scheme of the 2675 bp *Ssp*I–*Hin*dIII fragment containing the rubredoxin gene

The sequenced fragment is shown to be part of the cloned 3.9 kb *Bgl*II–*Hin*dIII fragment (insert of pCPRD1) and to include the 476 bp *Ssp*I–*Ssp*I fragment (insert of pCPRD2) used to improve the expression of the rubredoxin gene in *E. coli* (see the text). Rd, rubredoxin gene.

2). As previously reported for *C. pasteurianum* DNA (Tonomura *et al.*, 1965), the G+C content was found to be very low (30%), but slightly higher in coding regions (33–35%) (Graves *et al.*, 1985; Meyer & Gagnon, 1991).

ORF1, which is nearest to the *Ssp*I site, is preceded by a potential ribosome-binding site (GGTGGTG) and encodes a 308-residue protein with a calculated molecular mass of 32400 Da. This putative gene product is similar to a number of NAD(P)-dependent oxidoreductases, and in particular it shares 35% identical residues with *E. coli* thioredoxin reductase (Russel & Model, 1988) (Fig. 3). It also resembles the 207–521 segment of the F52a component of the alkyl hydroperoxide reductase system of *Salmonella typhimurium* (Tartaglia *et al.*, 1990) (Fig. 3). ORF1 is followed (with only an interval of one base) by a less likely (on the basis of codon usage) open reading frame (ORF2) which is preceded by a ribosome-binding site (AGGAGGT). The product of ORF2 would be a 78-residue protein which displays no significant similarity to any of the proteins of the NBRF-PIR (release 26) database.

Another open reading frame, ORF3, is located 288 bp downstream of ORF1. It is preceded by a ribosome-binding site (AGGAGG) and followed by a transcription termination signal consisting of a 20 bp stem with only one intervening base (Fig. 2). ORF3 potentially encodes a 178-residue protein (molecular mass 20037 Da) which bears similarity to the C-22 component of the alkyl hydroperoxide reductase from *S. typhimurium* (Tartaglia *et al.*, 1990) and to the product of a mouse gene, MER5, reported to be involved in erythroleukaemia cell differentiation (Nemoto *et al.*, 1990) (Fig. 4).

The rubredoxin-encoding gene starts 214 bases downstream from the termination codon of ORF3. It is preceded by a ribosome-binding sequence (GGAGG) and followed by a stem–loop transcription termination signal (Fig. 2). Several potential promoter sequences, consistent with known clostridial transcriptional signals (Young *et al.*, 1989), may be found upstream of the rubredoxin gene. The most likely ones, on the basis of rubredoxin expression experiments in *E. coli* (see below), are those nearest to the gene, at positions 2103–2108 for the −35 region and 2125–2130 for the −10 region (Fig. 2). Preliminary studies of the rubredoxin transcript indicate a size of approx. 230 bases, which is consistent with the putative transcription initiation and termination signals mentioned above.

The amino acid sequence deduced from the rubredoxin gene shows three differences from the one previously determined by Edman degradation (McCarthy, 1972; Watenpaugh *et al.*, 1973): D-14, D-22 and E-48 in the latter sequence are encoded by codons of the corresponding amides. Several explanations may rationalize these discrepancies, including sequencing errors and deamidation *in vivo* or *in vitro*. We therefore checked the questionable residues by amino acid sequencing. *C. pasteurianum* rubredoxin was carboxymethylated and treated with CNBr in order to remove the formylated *N*-terminal methionine. Subsequent automated Edman degradation yielded residues 2–33, all in perfect agreement with the gene sequence, including N-14 and N-22. In addition, carboxymethylated rubredoxin was digested with chymotrypsin, and the peptide mixture was run through a reversed-phase C_8 (Aquapore RP300; Brownlee Laboratories) h.p.l.c. column (Meyer *et al.*, 1990). Several peptides were sequenced, and in particular the one extending from residues 12 to 30, which confirmed N-14 and N-22, and the *C*-terminal one starting at C-42, which confirmed Q-48. Thus, the experimentally derived protein sequence is identical with that deduced from the gene sequence. A further argument supporting the presence of asparagine in position 14 was derived from the high-resolution X-ray structure of the protein (Watenpaugh *et al.*, 1979): the values of the B factors of the atoms of residue 14 allow the inference that the two δ atoms of this residue are not identical, i.e. they consist of one nitrogen and one oxygen, instead of two oxygens (L. Sieker, personal communication).

Rubredoxin expression in *E. coli*

Rubredoxin expression was first monitored in *E. coli* JM109 cells transformed with the pCPRD1 plasmid. Cells grown aerobically to $A_{600} = 1$ in LB medium were supplemented with [^{35}S]cysteine and analysed by SDS/PAGE. Rubredoxin, which co-migrated with the front, was detected on stained gels and in autoradiographs (Fig. 5), by comparison with the pure protein isolated from *C. pasteurianum*. The amounts of rubredoxin detected were not increased on addition of IPTG (Fig. 5, lanes 1 and 2). This indicated that the rubredoxin gene in pCPRD1 is not under the control of the *lac* promoter, but that its transcriptional and translational signals are efficient in *E. coli*. The putative products of ORF1 and ORF3 were not clearly visible on the background (Fig. 5).

In an attempt to improve the expression of the rubredoxin gene, we constructed a new plasmid containing only this gene. The 476 bp *Ssp*I–*Ssp*I fragment encompassing nucleotides 2090–2565 of the sequenced *Ssp*I–*Hin*dIII fragment (Figs. 1 and 2) was excised from pCPRD1 and subcloned into the *Sma*I site of pUC18. *E. coli* JM109 cells were transformed with the ligation mixture. Two transformants were shown by sequence analysis to contain the insert in the same orientation, with the rubredoxin gene downstream of the *lac* promoter. One of these, the plasmid of which was designated pCPRD2, has been used in the experiments described hereafter. In the absence of IPTG, the rubredoxin gene carried by pCPRD2 was expressed in similar amounts to those found on expression with pCPRD1 (Fig. 5, lanes 1, 2 and 6). This suggests that the −10 and −35 sequences shown in Fig. 2 constitute the active promoter of the rubredoxin gene in *E. coli*. The addition of IPTG to the growth medium resulted in a significant increase in rubredoxin expression (Fig. 5, lane 7). This increased level of rubredoxin was detectable after 1 h, levelled off after 4 h and remained stable for at least 24 h (not shown). No change in rubredoxin expression was observed on varying the induction point along the exponential phase of growth. The

aatattggtaattttgtaaaaaagcaatgctgtgcatattgcatattttttttgttatattataataatta 72

atataagtaaaaggctatattaattaataatgattattgtttaatgctgttaatgtaaaaacatctttata 144

tattaaaatataaaggtagtctaatatataaattatat**GGTGGTG**tttatgacatgaaagaagagaagcagc 216

M K E E K Q

ORF1

ttgatttagttataatcggagcaggtcccgcgggacttacggcggcgatttatgcaataagagctaaactca 288

L D L V I I G A G P A G L T A A I Y A I R A K L

atactttagttttagaaaatgaacttgtaggaggacaaattagggaaacttatacagtagaaaatttcccgg 360

N T L V L E N E L V G G Q I R E T Y T V E N F P

gatttaatgttatatctggtgctgatttggcagataaaatggaagagcatgcagcaagcattggagtaaata 432

G F N V I S G A D L A D K M E E H A A S I G V N

tagatcaatttagtaatatagaaaaaattaaattatctgatgatgaaaaaattatagaaacagaggatgtaa 504

I D Q F S N I E K I K L S D D E K I I E T E D V

tatataaagttaaggcgctgataatagctacaggagccaaaagcagaagactgcctattccagaagaagaaa 576

I Y K V K A L I I A T G A K S R R L P I P E E E

aactccatggaaaagttattcattattgtgagttatgtgatggagctctataccaaggaaaggatttggttg 648

K L H G K V I H Y C E L C D G A L Y Q G K D L V

ttgtgggcggtggaaattctgctgtagaagcagctatattcttaactaaatatgctagaaatattactatag 720

V V G G G N S A V E A A I F L T K Y A R N I T I

ttcatcaatttgactatcttcaagcacaaaagtactcacaggatgaattgtttaaacataaaaatgtaaaaa 792

V H Q F D Y L Q A Q K Y S Q D E L F K H K N V K

taatatgggatagtgagataagaaatattgttggagaaaatgagattgaaaaaatagtagtagaaaatgtaa 864

I I W D S E I R N I V G E N E I E K I V V E N V

agacaaaacagaaaacagaattgaaagcagatggagttttgtttatattggttatgagcctaagacggaac 936

K T K Q K T E L K A D G V F V Y I G Y E P K T E

tctttaaagattctattaatataaataaatggggatatatagaaaccgatgaaaatatggaaaccaatataa 1008

L F K D S I N I N K W G Y I E T D E N M E T N I

agggagtttttgctgctggagatgtaaggtctaaacttataagacagttaactacagctgttagtgatggca 1080

K G V F A A G D V R S K L I R Q L T T A V S D G

ctgttgctgcactaatggctgaaaaatatat**AGGAGGT**aaataatatgattaaaatatattccacaccaaca 1152

T V A A L M A E K Y I G G K * M I K I Y S T P T

ORF2

tgtccatggtgtaaaaagacaaaagaatatttaaaatcaaaaaatatagattttgttgatgtaaatgtagct 1224

C P W C K K T K E Y L K S K N I D F V D V N V A

gatgatatgaaggaaagagaagaaatgcgtagcttatccaagcaatcaggagtgccagttataaatatagat 1296

D D M K E R E E M R S L S K Q S G V P V I N I D

ggtaatataatagttggatttaacaaggcagaaatagataaactaatagagaaatagactactttttggtat 1368

G N I I V G F N K A E I D K L I E K *

ggtatatatggtatatgatgaacaaa**AGGAGG**agatttatggagagattagtgggaaagccagcaccagaa 1440

M E R L V G K P A P E

ORF3

tttgaaatgaaggctgtcaaaggtgatggaagaggatttactgaagttaaactgggagattataagggaaaa 1512

F E M K A V K G D G R G F T E V K L G D Y K G K

tggctagtaatgtttttttatccgttagatttacatttgtatgcccaactgaaattacaggctttagcaaa 1584

W L V M F F Y P L D F T F V C P T E I T G F S K

agggctgaagaatttagggatttaaaagctgaattattagcggtaagttgtgatagtcaatattctcatgag 1656

R A E E F R D L K A E L L A V S C D S Q Y S H E

acttggattaatcaagatataaaacagggggggacttggtaaaattaatttccaattgcctccgataaaact 1728

T W I N Q D I K Q G G L G K I N F P I A S D K T

actgaagtgagtacaaaatatggaatacagatagaagaagaaggtatttctctaagaggactatttataatt 1800

T E V S T K Y G I Q I E E E G I S L R G L F I I

gatcctgaaggaatagttaggtattctgtagtgcatgatcttaatgtaggaagaagtgttgatgaaactctt 1872

D P E G I V R Y S V V H D L N V G R S V D E T L

cgtgtattaaaggcatttcaaactggtggaatgtgtgctttagattggcatgaaggagatgacaacttataa 1944

R V L K A F Q T G G M C A L D W H E G D D N L *

tttatggggatttgatcttatatgactagactaaattttatt**attaaaacttgaaatatata**c**tatatattt** 2016

caagttttaattgaagttcatattgcataaatttttgttaaataagtatatttttataaaataaggtagaaa 2088

tatttacttaattg***TTGTAA***tataaaataaatacat***TAAAAT***ttattaaggtatt**GGAGG**agtttttatga 2160

Rd M

agaagtatacatgtacagtatgtggatatatttataatcccgaagatggagacccagataatggagtaaatc 2232

K K Y T C T V C G Y I Y N P E D G D P D N G V N

caggaacagattttaaagatattccagatgattgggtatgtcctttgtgtggagtaggaaaagatcaatttg 2304

P G T D F K D I P D D W V C P L C G V G K D Q F

aagaagtagaagaataggata**aattaaagactgggag**agaatttg**cttccagtctttaatt**atgattcacta 2376

E E V E E *

tatttgtatatctacattatataatcttagccaagtatctacttgaataagataagctattaattgaggtcc 2448

tgtcattaattgaccataccaaggtttttataggatttacctccagttctcacaatctctcttactacagg 2520

aacatcgatcaattgaagtataggagaattaggatcatttaaaatattttcaagccattttgtactatttt 2592

agtatatccaggattaaaggttttaggataaggactctttttctatagacaatatcatcaggaagtattcc 2664

ctttaaagctt 2675

```
      1                 20                  40                   60
ORF1  MKEEKQLDLVIIGAGPAGLTAAIYAIRAKLNTLVLENELVGGQIRETYTVENFPG.FNVIS
      *   *   * * * **** *** ** ** *          ***   *  *** **  *
TXRD  MGTTKHSKLLILGSGPAGYTAAVYAARANLQPVLITGMEKGGQLTTTTEVENWPGDPNDLT
      208      ** ****** **** **      *  * * ***   *   **    *
F52A  LNKRDAYDVLIVGSGPAGAAAAVYSARKGIRTGLM.GREFGGQVLDTVDIENYISVPKT.E

                        80                     100
ORF1  GADLADKMEEHAASIGVNIDQFSNIEKIKLSDDEKIIE....TEDVIYKVKALIIATGAKS
      *  *   * ***      *  *  * *  *                 *   ********
TXRD  GPLLMERMHEHATKFETEII.FDHINKVDLQNRPFRLN....GDNGEYTCDALIIATGASA
      *  *      *        *      *         *                ******
F52A  GQKLAGALKAHVSDYDVDVIDSQSASKLVPAATEGGLHQIETASGAVLKARSIIIATGAKW

      120                 140                 160
ORF1  RRLPIPEEEKLHGKVIHYCELCDGALYQGKDLVVVGGGNSAVEAAIFLTKYARNITIVHQF
      * *  * **   *     *  ***  *      * **** *** *  *   *      *
TXRD  RYLGLPSEEAFKGRGVSACATCDGFFYRNQKVAVIGGGNTAVEEALYLSNIASEVHLIHRR
      *    * *      **  *  ***       ********  ** *  *  *   * *
F52A  RNMNVPGEDQYRTKGVTYCPHCDGPLFKGKRVAVIGGGNSGVEAAIDLAGIVEHVTLLEFA

     180                  200                   220
ORF1  DYLQAQKYSQDELF.KHKNVKIIWDSE..IRNIVGEN.EIEKIVVENVKTKQKTE.LKADG
      *   * *     *  *  *  **           *                   * *   *
TXRD  DGFRAEKILIKRLMDKVENGNIILHTNRTLEEVTGDQMGVTGVRLRDTQNSDNIESLDVAG
          *   *         *  ***       ** **   * *   **   *  * *   **
F52A  PEMKADQVLQDKVR.SLKNVDIILNAQ..TTEVKGDGSKVVGLEYRD.RVSGDIHSVALAG

         240                 260                      280
ORF1  VFVYIGYEPKTELFKDSININKWGYIETDENM.....ETNIKGVFAAGDVRSKLIRQLTTA
       ** **  * *  *         ***            * * ********     **  *
TXRD  LFVAIGHSPNTAIFEGQLELEN.GYIKVQSGIHGNATQTSIPGVFAAGDVMDHJYRQAITS
       ** **  ***   ** **    * *            **  *******     * * *
F52A  IFVQIGLLPNTHWLEGALERNRMGEIIIDAKC.....ETSVKGVFAAGDCTTVPYKQIIIA

             300     308
ORF1  VSDGTVAALMAEKYIGGK
         *  *** ** *  *
TXRD  AGTGCMAALDAERYLDGLADAK
       * *  * * *  **     *
F52A  TGEGAKASLSAFDYLIRTKIA
```

Fig. 3. Comparison of the sequence of ORF1 with those of the thioredoxin reductase (TXRD) from *E. coli* (Russel & Model, 1988) and the *C*-terminal part (starting at residue 208) of the F52a component of *S. typhimurium* alkyl hydroperoxide reductase (Tartaglia *et al.*, 1990)

The alignments were carried out using MULTALIN (Corpet, 1988). Matching residues are marked with stars. The numbers refer to the ORF1 sequence only, the designated residues being below the last digit of each number. Several regions of strong similarity are apparent, some of which, namely 10–28, 149–165 and 267–278, correspond to nucleotide-binding domains (Tartaglia *et al.*, 1990; Eggink *et al.*, 1990). The redox-active disulphide bridge is formed by C-136 and C-139.

```
     20        30                  50                  70
ORF3  DGRGFTEVKLGDYKGKWLVMFFYPLDFTFVCPTEITGFSKRAEEFRDLKAELLAVSCDSQ
     79   * *  * * *** ** ***************  **  * ** *   *  *** **
MER5  VNGEFKELSLDDFKGKYLVLFFYPLDFTFVCPTEIVAFSDKANEFHDVNCEVVAVSVDSH
     17**** *    *  *   * **** *********     *   *       *  ** * *
C22   KNGEFIEVTEKDTEGRWSVFFFYPADFTFVCPTELGDVADHYEELQKLGVDVYSVSTDTH

               90                  110                  130
ORF3  YSHETWINQDIKQGGLGKINFPIASDKTTEVSTKYGIQIEEEGIS.LRGLFIIDPEGIVR
       **  ***   * ****  *    ** *   *  **   *  **  ********* * *
MER5  FSHLAWINTPRKNGGLGHMNITLLSDITKQISRDYGVLLESAGIA.LRGLFIIDPNGVVK
      * * **                   * *    *      *  * *  * *
C22   FTHKAW...HSSSETIAKIKYAMIGDPTGALTRNFDNMREDEGLADVRPLLLTRRVSSRR

               150                  170
ORF3  YSVVHDLNVGRSVDETLRVLKAFQ....TGGMCALDWHEGDDNL
         * **  ***** ****  ****     *  *   *
MER5  HLSVNDLPVGRSVEETLRLVKAFQFVETHGEVCPANWTPESPTIKPSPTASKEYFEKVHQ
                        *        *       *
C22   SKLPLKVSAVTRLTCCVKLKQHSTLPLTQAKYARRNGKKANDSGSILRPGR
```

Fig. 4. Comparison of the sequence of ORF3 (starting at residue 20) with those of the protein encoded by MER5 (Nemoto *et al.*, 1990) and the C-22 component of the alkyl hydroperoxide reductase from *S. typhimurium* (Tartaglia *et al.*, 1990)

The alignments were carried out using MULTALIN (Corpet, 1988). Matching residues are marked with stars. The numbers refer to the ORF3 sequence only, the designated residues being below the last digit of each number.

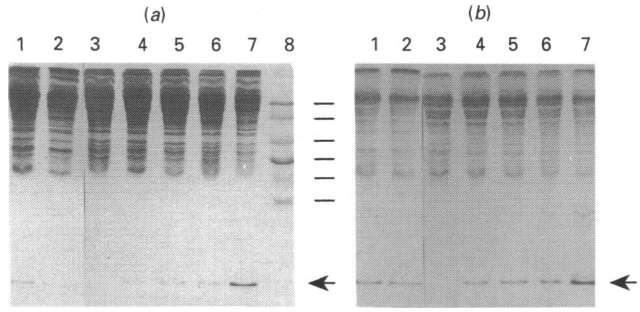

Fig. 5. Analysis of rubredoxin expression in *E. coli* by SDS/PAGE

E. coli JM109 cells containing the plasmids indicated below were grown in LB to $A_{600} = 1$. At this stage [^{35}S]cysteine was added to all cultures at a final concentration of 5 μCi/ml. At the same time, IPTG (1 mM final concentration) was added as indicated. After overnight incubation, 1 ml of cell suspension was centrifuged, and the pellet was treated at 95 °C for 5 min with 0.1 ml of lysis buffer (Laemmli, 1970). Cell debris was eliminated by centrifugation and 5 μl of supernatant was loaded on the gel. The acrylamide concentration was 10 % (w/v) in the stacking gel and 20 % (w/v) in the separating gel. (*a*) Coomassie Blue staining. (*b*) Autoradiography. Lane 1, pCPRD1; lane 2, pCPRD1 + IPTG; lane 3, pUC19 + IPTG; lane 4, pCPRD3; lane 5, pCPRD3 + IPTG; lane 6, pCPRD2; lane 7, pCPRD2 + IPTG; lane 8, molecular-mass markers (indicated with lines: ovalbumin, 46 kDa; carbonic anhydrase, 30 kDa; trypsin inhibitor, 21.5 kDa; lysozyme, 14.3 kDa; aprotinin, 6.5 kDa; insulin a, 3.4 kDa).

dependence of rubredoxin expression on the presence of the *lac* promoter upstream of the gene in pCPRD2 was confirmed by cloning the 476 bp *Ssp*I–*Ssp*I fragment in pUC19: the insert containing the rubredoxin gene was excised from pCPRD2 by a *Hin*dIII–*Eco*RI double digestion, and ligated into the corresponding sites of pUC19. In the resulting plasmid, pCPRD3, the rubredoxin gene and the *lac* promoter are oriented in opposite directions. The level of rubredoxin expression was indeed found to be insensitive to IPTG, and was similar to that observed with pCPRD2 in the absence of IPTG (Fig. 5, lanes 4 and 5).

These experiments show that the rubredoxin polypeptide is synthesized in *E. coli*, but produce no information regarding the assembly of holorubredoxin. We therefore undertook the extraction and purification (see the Materials and methods section) of rubredoxin from *E. coli* JM109 cells transformed with pCPRD2. Cells grown overnight in LB medium yielded 1–2 mg of rubredoxin/l of culture. The addition of IPTG (1 mM) in mid-log phase ($A_{600} = 1$) resulted in a doubling of rubredoxin production. The characteristic red colour of rubredoxin was already detectable in the centrifuged cell extract. The core and tail fractions of the main peak eluted from the Sephadex G-50 column (second purification step) contained mostly rubredoxin ($A_{490}/A_{280} > 0.30$). When these fractions were run through an anion-exchange h.p.l.c. column (see the Materials and methods section), two peaks were eluted: the first one was pink and proved to be pure rubredoxin ($A_{490}/A_{280} = 0.43$). The second one was colourless and was shown to be aporubredoxin by its migration, which was identical with that of rubredoxin on SDS/polyacrylamide gels, and by its conversion into holorubredoxin on incubation under argon in the presence of dithiothreitol and ferric ion. The amounts of holo- and apo-rubredoxin were found to be approximately equal. In attempts to increase the

Fig. 2. Nucleotide sequence of the 2675 bp *Ssp*I–*Hin*dIII fragment

The translated amino acid sequences of the open reading frames (ORF) are written below the corresponding DNA-coding regions. The sequences corresponding to ribosome-binding sites of mRNA are shown in underlined bold capitals. The putative promoter regions of the rubredoxin gene are shown in bold italic capitals. The transcription termination signals are underlined with arrows and a dot to mark the centre of symmetry.

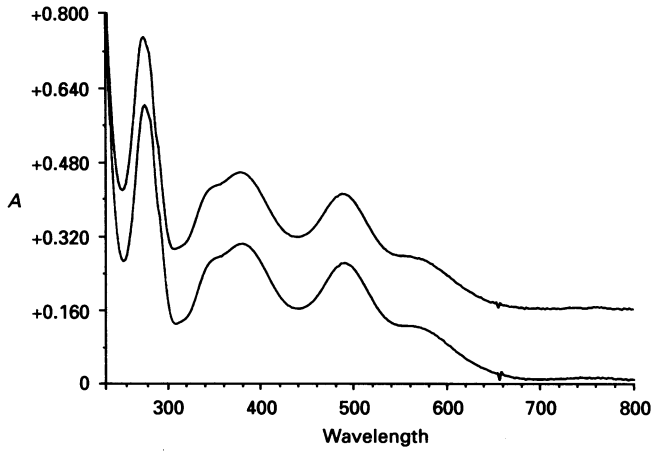

Fig. 6. U.v.–visible-absorption spectra of *C. pasteurianum* rubredoxin as synthesized in *C. pasteurianum* (upper spectrum) and in *E. coli* (lower spectrum)

For the sake of clarity, an offset of +0.15 absorption unit has been applied to the spectrum of the native *C. pasteurianum* rubredoxin. Both proteins (concentration 1.6 mg/ml) were in 20 mM-Tris/HCl/0.4 M-NaCl, pH 8.0. The spectra were recorded on a Hewlett–Packard 8452 diode array spectrophotometer.

yield of holorubredoxin, growth media were supplemented with $FeCl_3$ (10 or 100 μM), but neither the total amount of rubredoxin, nor the ratio of holo- to apo-protein were modified. Thus the concentration of iron in the LB medium does not limit the production of holorubredoxin by *E. coli*. Assuming that apo-rubredoxin was present in the crude extract, we tried to convert it into holorubredoxin: the extract was incubated under argon with 0.5 mM-dithiothreitol for several hours at 4 °C, then supplemented with 20 μM-$FeCl_3$, and incubated for 1 h at room temperature. Subsequent purification of rubredoxin as described above revealed no increase in the ratio of holo- to apo-rubredoxin. It thus remains to be established if the presence of aporubredoxin arises from the inability of *E. coli* to provide all the protein molecules with iron, or from the loss of the active site during purification.

The u.v.–visible-absorption spectrum of the rubredoxin expressed in *E. coli* is identical with that of native rubredoxin from *C. pasteurianum* (Fig. 6). The same holds true for the e.p.r. spectra (not shown). These data demonstrate that the two proteins have identical chromophores. Both proteins have been analysed by electrospray m.s., a technique that allows measurement of the molecular mass of proteins, including rubredoxins (Meyer *et al.*, 1990). The mass of the rubredoxin expressed in *E. coli* was found to be smaller than that of the native protein by 28 mass units, a difference which suggested the removal of the *N*-terminal blocking formyl group. *N*-Terminal sequencing of recombinant rubredoxin confirmed that its *N*-terminus was unblocked; the average sequencing yield for the three *N*-terminal residues (M,K,K) was 40%. In contrast, the sequencing yield of the rubredoxin purified from *C. pasteurianum* was approx. 5%, thus confirming a previous report that the *N*-terminus was blocked (McCarthy, 1972).

DISCUSSION

The structure of the rubredoxin-encoding gene of *C. pasteurianum* shows that in this bacterium the polypeptide is synthesized in its final size, as in *D. vulgaris* (Voordouw, 1988). A possible promoter region occurs upstream from the gene, and a transcription termination signal is found shortly downstream (Fig. 2). These features are consistent with preliminary studies on the rubredoxin transcript, which indicate a size of approx. 230 bases. Thus the rubredoxin gene probably constitutes an independent operon. Additional transcript-mapping experiments are required to confirm this hypothesis.

The situation of the rubredoxin gene in the *C. pasteurianum* genome is quite unlike the one observed in *P. oleovorans*, where the rubredoxin gene belongs to a multigene operon encoding enzymes involved in alkane hydroxylation (Eggink *et al.*, 1987, 1990; Kok *et al.*, 1989), or in *D. vulgaris*, where the rubredoxin gene is adjacent to a gene encoding another metalloprotein, desulphoferrodoxin (Brumlik & Voordouw, 1989; Moura *et al.*, 1990). The aerobe *P. oleovorans* remains the only organism in which the function of rubredoxin has been well characterized (Eggink *et al.*, 1990). In anaerobes, despite their great structural similarities, rubredoxins may well fulfil different functions in the diverse bacteria that contain them. Various means of identifying possible redox partners of rubredoxin in *C. pasteurianum* are possible; one of these consists of characterizing and quantifying the transcripts of the rubredoxin gene under various conditions. The 476 bp *Ssp*I–*Ssp*I fragment cloned in pCPRD2 will therefore constitute a useful RNA probe.

With the present data, no functional connection between rubredoxin and the putative translational products of ORF1 and ORF3 occurring on the sequenced fragment can be inferred. In view of its strong similarity to *E. coli* thioredoxin reductase (Fig. 3), the product of ORF1 might be the counterpart enzyme of *C. pasteurianum*: a thioredoxin–thioredoxin reductase system has indeed been isolated from this bacterium (Hammel *et al.*, 1983). In a sequence alignment of the putative product of ORF1 and of *E. coli* thioredoxin reductase, particularly strong similarities are observed around the nucleotide-binding regions (Tartaglia *et al.*, 1990; Eggink *et al.*, 1990; McKie & Douglas, 1991), and around the two conserved redox-active cysteines of thioredoxin reductase (Fig. 3). Since the thioredoxin gene of *C. pasteurianum* does not occur on the 3.9 kb *Bgl*II–*Hin*dIII fragment, it must be separated from the reductase gene by more than 1.5 kb. However, the presence of the thioredoxin-encoding gene in close vicinity to the reductase-encoding one is not mandatory, as in *E. coli* the two corresponding genes have been mapped in very distant loci (Haller & Fuchs, 1984; Lim *et al.*, 1985).

The putative product of ORF3 exhibits greatest similarity with the product of the mouse MER5 gene which has been proposed to be involved in leukaemia cell differentiation (Nemoto *et al.*, 1990). The latter protein is larger (257 residues) than the product of ORF3, but the similarity is so high (53% identity) in the approx. 160-residue overlap that the two proteins might have related functions. An alternative proposal can be made, which is less satisfactory with regard to sequence similarities, but which would provide a function for the products of both ORF1 and ORF3; they are similar to the *ahpF* and *ahpC* genes, respectively, that encode the two components of the alkyl hydroperoxide reductase of *S. typhimurium* (Figs. 3 and 4) (Tartaglia *et al.*, 1990). In the *C. pasteurianum* genome, the order of occurrence of the former genes is opposite to that of the latter in *S. typhimurium*.

The rubredoxin sequence as translated from the gene is very similar to the original sequence determination by Edman degradation (McCarthy, 1972; Watenpaugh *et al.*, 1973). Three minor discrepancies, amides in place of the corresponding amino acids, have been confirmed by resequencing the protein purified from *C. pasteurianum*. The first change, D-14 ⇒ N, is somewhat surprising, since aspartate was found to occur in this position in all rubredoxin sequences known to date (Meyer *et al.*, 1990; Fig. 7). The second change, D-22 ⇒ N, is consistent with the available sequence data, since, although residue 22 is altogether a variable site, amides occur at this position in all clostridial rubredoxins. Similarly, the last modification, E-48 ⇒ Q, is in line with the

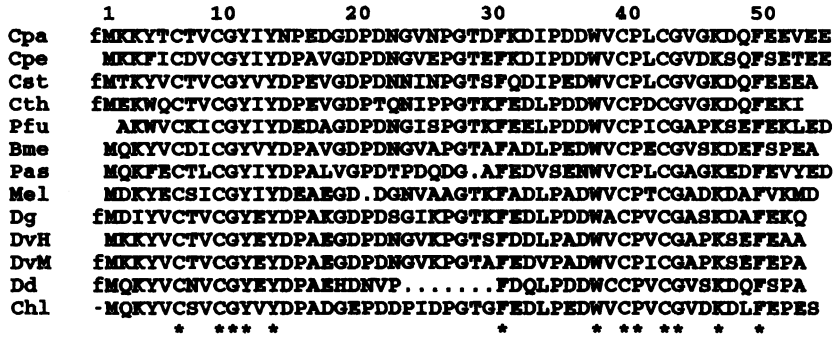

Fig. 7. Alignment of the presently known rubredoxin sequences

The primary sources for the sequences from *Clostridium perfringens* (Cpe), *Clostridium thermosaccharolyticum* (Cth), *Butyribacterium methylotrophicum* (Bme), *Peptostreptococcus asaccharolyticus* (Pas), *Megasphaera elsdenii* (Mel), *Desulfovibrio gigas* (Dg), *Desulfovibrio vulgaris* Hildenborough (DvH), *Desulfovibrio desulfuricans* (Dd) and *Chlorobium limicola* (Chl) are to be found in Meyer *et al.* (1990). The remaining ones are *C. pasteurianum* (Cpa, this work), *Clostridium sticklandii* (Cst; J. Meyer, J. Gagnon & J.-M. Moulis, unpublished work), *D. vulgaris* Miyazaki (DvM; Shimizu *et al.*, 1989), and *Pyrococcus furiosus* (Pfu; Blake *et al.*, 1991). The hyphen in front of the Chl sequence is to indicate that the status of the *N*-terminal methionine (blocked or unblocked) was not clearly stated in the original reference.

presence of glutamine at this position in all clostridial rubredoxins (Meyer *et al.*, 1990; Fig. 7). The 13 rubredoxin sequences known to date have been aligned in Fig. 7. The totally conserved residues now amount to 13, which include the four cysteine ligands of the iron atom, three prolines or glycines involved in formation of the cysteine-containing loops, and five of the six aromatic residues that constitute the hydrophobic core of the protein. Most of these residues are presumably essential to maintain the architecture of the rubredoxin molecule and the redox properties of its iron ion. In contrast, among the charged residues on the surface of the protein, only K-46 is totally conserved.

C. pasteurianum rubredoxin as expressed in *E. coli* possesses an iron chromophore identical with the one present in the native protein (Fig. 6). The ability of *E. coli* to assemble the active site of this metalloprotein is not unexpected, in view of its reported efficiency in expressing foreign proteins possessing more elaborate iron–sulphur clusters (Makaroff *et al.*, 1983; Coghlan & Vickery, 1989; Böhme and Haselkorn, 1989; Baur *et al.*, 1990; Grabau *et al.*, 1991; V. Davasse & J.-M. Moulis, unpublished work). *E. coli* may use an indigenous enzyme system to insert iron at the active site of *C. pasteurianum* rubredoxin, although rubredoxin-like proteins have yet to be found in wild-type *E. coli*. Alternatively, it is not unreasonable to consider a spontaneous assembly of holorubredoxin *in vivo*. We have found, on the basis of *N*-terminal sequencing and m.s. analysis (see the Results section), a minor difference between the products of the rubredoxin gene in *C. pasteurianum* and that in *E. coli*; in the latter organism, rubredoxin undergoes removal of the formyl group from the *N*-terminal methionine, whereas in *C. pasteurianum* the *N*-terminus remains blocked.

The level of expression presently obtained for *C. pasteurianum* rubredoxin in *E. coli* represents a considerable improvement over rubredoxin expression in *C. pasteurianum*; in the latter case yields are lower than 0.1 mg/l of culture and are not very reproducible, for as yet unknown reasons (I. Mathieu, J. Meyer & J.-M. Moulis, unpublished work). In contrast, we have succeeded in obtaining yields consistently higher than 2 mg/l of culture from *E. coli* transformed with pCPRD2. This, in addition to providing an experimentally feasible system for the controlled mutagenesis of rubredoxin, is of considerable interest for the production of the large amounts of native rubredoxin required for structural and functional investigations.

We thank J. Gagnon and N. Scherrer for amino acid sequencing, G. Hegy and A. Van Dorsselaer for performing the m.s. analyses, L. Sieker for discussions on the structure of rubredoxins, J. Gaillard for recording e.p.r. spectra, and P. Richaud for the synthesis of oligonucleotides and for helpful advice on DNA sequencing.

REFERENCES

Adman, E. T., Sieker, L. C. & Jensen, L. H. (1991) J. Mol. Biol. **217**, 337–352
Ballongue, J., Amine, J., Masion, E., Petitdemange, H. & Gay, R. (1986) Biochimie **68**, 575–580
Baur, J. R., Graves, M. C., Feinberg, B. A. & Ragsdale, S. W. (1990) BioFactors **2**, 197–203
Blake, P. R., Park, J.-B., Bryant, F. O., Aono, S., Magnuson, J. K., Eccleston, E., Howard, J. B., Summers, M. F. & Adams, M. W. W. (1991) Biochemistry **30**, 10885–10895
Böhme, H. & Haselkorn, R. (1989) Plant Mol. Biol. **12**, 667–672
Brumlik, M. J. & Voordouw, G. (1989) J. Bacteriol. **171**, 4996–5004
Chen, K. C.-K., Chen, J.-S. & Johnson, J. L. (1986) J. Bacteriol. **166**, 162–172
Coghlan, V. M. & Vickery, L. E. (1989) Proc. Natl. Acad. Sci. U.S.A. **86**, 835–839
Corpet, F. (1988) Nucleic Acids Res. **16**, 10881–10890
Eggink, G., Lageveen, R. G., Altenburg, B. & Witholt, B. (1987) J. Biol. Chem. **262**, 17712–17718
Eggink, G., Engel, H., Vriend, G., Terpstra, P. & Witholt, B. (1990) J. Mol. Biol. **212**, 135–142
Frey, M., Sieker, L. C., Payan, F., Haser, R., Bruschi, M., Pepe, G. & LeGall, J. (1987) J. Mol. Biol. **197**, 525–541
Gebhard, M. S., Deaton, J. C., Koch, S. A., Millar, M. & Solomon, E. I. (1990) J. Am. Chem. Soc. **112**, 2217–2231
Grabau, C., Schatt, E., Jouanneau, Y. & Vignais, P. M. (1991) J. Biol. Chem. **266**, 3294–3299
Graves, M. C., Mullenbach, G. T. & Rabinowitz, J. C. (1985) Proc. Natl. Acad. Sci. U.S.A. **82**, 1653–1657
Haller, B. L. & Fuchs, J. A. (1984) J. Bacteriol. **159**, 1060–1062
Hammel, K. E., Cornwell, K. L. & Buchanan, B. B. (1983) Proc. Natl. Acad. Sci. U.S.A. **80**, 3681–3685
Hinton, S. M., Slaughter, C., Eisner, W. & Fisher, T. (1987) Gene **54**, 211–219
Kieser, T. (1984) Plasmid **12**, 19–36
Kok, M., Oldenhuis, R., van der Linden, M. P. G., Raatjes, P., Kingma, J., van Lelyveld, P. H. & Witholt, B. (1989) J. Biol. Chem. **264**, 5435–5441
Laemmli, U. K. (1970) Nature (London) **227**, 680–685
LeGall, J. (1968) Ann. Inst. Pasteur **114**, 109–115
Lim, C.-J., Haller, B. & Fuchs, J. A. (1985) J. Bacteriol. **161**, 799–802

Lovenberg, W. & Sobel, B. E. (1965) Proc. Natl. Acad. Sci. U.S.A. **54**, 193–199
Makaroff, C. A., Zalkin, H., Switzer, R. L. & Vollmer, S. J. (1983) J Biol. Chem. **258**, 10586–10593
McCarthy, K. F. (1972) Ph.D. Thesis, George Washington University, Washington, DC
McKie, J. H. & Douglas, K. T. (1991) FEBS Lett. **279**, 5–8
Meyer, J. & Gagnon, J. (1991) Biochemistry **30**, 9697–9704
Meyer, J., Gagnon, J., Sieker, L. C., Van Dorsselaer, A. & Moulis, J.-M. (1990) Biochem. J. **271**, 839–841
Moulis, J.-M. & Meyer, J. (1982) Biochemistry **21**, 4762–4771
Moura, I., Tavares, P., Moura, J. J. G., Ravi, N., Huynh, B. H., Liu, M.-Y. & LeGall, J. (1990) J. Biol. Chem. **265**, 21596–21602
Nemoto, Y., Yamamoto, T., Takada, S., Matsui, Y. & Obinata, M. (1990) Gene **91**, 261–265
Rabinowitz, J. C. (1972) Methods Enzymol. **24B**, 431–446
Russel, M. & Model, P. (1988) J. Biol. Chem. **263**, 9015–9019
Saito, H. & Miura, K.-I. (1963) Biochim. Biophys. Acta **72**, 619–629
Sanger, F., Nicklen, S. & Coulson, A. R. (1977) Proc. Natl. Acad. Sci. U.S.A. **74**, 5463–5467
Seki, S., Ikeda, A. & Ishimoto, M. (1988) J. Biochem. (Tokyo) **103**, 583–584
Shimizu, F., Ogata, M., Yagi, T., Wakabayashi, S. & Matsubara, H. (1989) Biochimie **71**, 1171–1177
Stadtman, T. C. (1965) in Non-Heme Iron Proteins: Role in Energy Conversion (San Pietro, E., ed.), pp. 439–445, Antioch Press, Yellow Springs, OH
Stenkamp, R. E., Sieker, L. C. & Jensen, L. H. (1990) Proteins: Struct. Funct. Genet. **8**, 352–364
Tartaglia, L. A., Storz, G., Brodsky, M. H., Lai, A. & Ames, B. N. (1990) J. Biol. Chem. **265**, 10535–10540
Tonomura, B. I., Malkin, R. & Rabinowitz, J. C. (1965) J. Bacteriol. **89**, 1438–1439
Voordouw, G. (1988) Gene **69**, 75–83
Wang, S.-Z., Chen, J.-S. & Johnson, J. L. (1988) Nucleic Acids. Res. **16**, 439–454
Wang, S.-Z., Chen, J.-S. & Johnson, J. L. (1990) Biochem. Biophys. Res. Commun. **169**, 1122–1128
Watenpaugh, K. D., Sieker, L. C., Herriott, J. R. & Jensen, L. H. (1973) Acta Crystallogr. **B29**, 943–956
Watenpaugh, K. D., Sieker, L. C. & Jensen, L. H. (1979) J. Mol. Biol. **131**, 509–522
Young, M., Minton, N. P. & Staudenbauer, W. L. (1989) FEMS Microbiol. Rev. **63**, 301–326

Received 21 October 1991/3 December 1991; accepted 12 December 1991