The pseudoazurin gene from *Thiosphaera pantotropha*: analysis of upstream putative regulatory sequences and overexpression in *Escherichia coli*

Yun-Chung LEUNG^{||}, Christopher CHAN[†], John S. READER[¶], Anthony C. WILLIS[‡], Rob J. M. VAN SPANNING[§], Stuart J. FERGUSON^{*†**} and Sheena E. RADFORD^{¶**}

^{*}Oxford Centre for Molecular Sciences, New Chemistry Laboratory, [†]Department of Biochemistry, and [‡]MRC Immunochemistry Unit, Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, U.K., and [§]Department of Microbiology, Biological Laboratory, Vrije Universiteit, De Boelelaan 1087, 1081 HV Amsterdam, The Netherlands

The pseudoazurin gene from *Thiosphaera pantotropha* has been cloned and sequenced. The deduced amino acid sequence showed that the protein contains an unusually alanine-rich signal peptide, 22 amino acid residues in length, which targets the protein to the periplasm. This pseudoazurin was expressed in large amounts in the periplasm of *Escherichia coli* when the gene with its native ribosome-binding site was placed downstream of the *lac* promoter. Removal of a putative hairpin-forming structure upstream

of the ribosome-binding site increased the yield of the purified protein to ~80 mg/l. The recombinant protein is indistinguishable from that purified from its natural host. A primer extension study indicated that the pseudoazurin structural gene (*pazS*) is under the control of the *Fnr/Nnr* regulatory system, but no promoter-binding sequence could be recognized. The amino acid sequence of pseudoazurin from *Paracoccus denitrificans* is also reported.

INTRODUCTION

Denitrification is an important biological process in which inorganic nitrate and nitrite are converted into gaseous nitrogen in a number of steps which involve reductases for nitrate, nitrite, nitric oxide and nitrous oxide. Several of these enzymes are supplied with electrons by copper proteins and/or *c*-type cytochromes [1]. Pseudoazurins are small copper-containing redox proteins present in some denitrifying bacteria. In their oxidized state, pseudoazurins and other cupredoxins are characterized by a strong peak in their absorption spectrum near 600 nm and a very small hyperfine coupling constant in their EPR spectrum [2]. These shuttle proteins transport electrons from various donors to acceptors within the bacterial periplasm. In some bacteria, the synthesis of pseudoazurin is enhanced under denitrifying conditions, and a role for pseudoazurin as an electron donor for soluble nitrous oxide reductase [3], copper-containing nitrite reductase (C. Chan and S. J. Ferguson, unpublished work), cytochrome *cd*₁ nitrite reductase [4,5] and membrane-bound nitric oxide reductase [5] has been proposed. It has been reported recently that pseudoazurin may be involved in a wide range of electron-transfer reactions with a number of structurally very different donors and acceptors, suggesting that the interaction between redox proteins is only pseudo-specific [6].

The pseudoazurin isolated from *Thiosphaera pantotropha* (renaming as *Paracoccus pantotrophus* is under consideration [7]; I. P. Thompson, S. J. Ferguson and S. C. Baker, unpublished work) contains 123 amino acids, with a molecular mass of 13 kDa [9]. The 2.5 Å crystal structure of the oxidized form of

this protein has recently been determined [6]. This revealed the unique observation that this pseudoazurin crystallizes as a dimer in the asymmetric unit, whereas all other pseudoazurins of known structure are monomeric [2]. Interestingly, the reduced form of the *T. pantotropha* pseudoazurin is monomeric, suggesting the intriguing possibility that a change in molecular mass might control the redox properties of the protein [6]. In common with other pseudoazurins, the fold of this polypeptide chain follows the eight-stranded Greek-key motif with a type I copper site. In contrast with the related plastocyanins, however, there are two α -helices located towards the C-terminus of the molecule [2].

Further understanding of the structure and roles of pseudoazurin from *T. pantotropha* will require substantial amounts of protein that can only be obtained by cloning and efficient expression of the *pazS* gene (yields of only 0.6 mg/l are possible with the naturally expressed protein [4]). The present paper reports studies that have achieved this aim, together with an analysis of the basis for expression of pseudoazurin in its natural host only under anaerobic conditions [10].

MATERIALS AND METHODS

Bacterial strains

T. pantotropha strain LMD 82.5 was used as the source of the pseudoazurin gene (*pazS*). *Escherichia coli* XL1-Blue [*endA*1, *hsdR*17 (rk[−], mk⁺), *supE*44, *thi*[−], λ [−], *recA*1, *gyrA*96, *relA*1 (*lac*), [F', *proAB*, *lacIq*, *lacZ*ΔM15, *Tn10Tc*]] was used as host for cloning and expression of *T. pantotropha pazS*. Cells were

Abbreviations used: RBS, ribosome-binding site; IPTG, isopropyl β -D-thiogalactoside.

^{||} Present address: Department of Anatomy, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong.

[¶] Present address: Department of Biochemistry and Molecular Biology, University of Leeds, Leeds LS2 9JT, U.K.

^{**} To whom correspondence should be addressed.

The nucleotide sequence data reported have been deposited in the EMBL/GenBank/DBJ Nucleotide Sequence Database under accession no. Z73141.

grown on $2 \times$ TY medium [11], supplemented with $100 \mu\text{g/ml}$ ampicillin, at 37°C under aerobic conditions. Recombinant DNA techniques were used according to standard protocols [11].

The uracil-containing template used for site-directed deletion mutagenesis [12] was prepared by using *E. coli* strain RZ1032 {HFr KL16 PO/45 [*lysA*-(16–21)], *dut-1*, *ung-1*, *thi-1*, *relA1*, *zbd-279::Tn10*, *supE44*}.

Oligonucleotides

Oligonucleotides were synthesized on an Applied Biosystems DNA Synthesizers model 380A. Positions of 4-fold base degeneracy were incorporated into the oligonucleotide by 'mixed-base' synthesis, adding all four nucleotide monomers at the coupling step.

PCR amplification

This was done using *Taq* DNA polymerase (Perkin Elmer-Cetus Instruments) and a MiniCycler (MJ Research) programmed for: step 1, denaturation at 95°C for 1 min; step 2, annealing at 37°C for 2 min; step 3, chain extension at 72°C for 2 min. Steps 1–3 were repeated 26 times and then the mixture was held at 4°C .

The PCR product was blunted by end-filling with the Klenow fragment of DNA polymerase I and was then inserted into the *Sma*I site of the phagemid pBluescript KS[−] (Stratagene) to yield plasmid pTL3.

DNA sequencing

Small-scale preparation of plasmid DNA was carried out using the modified mini alkaline-lysis/poly(ethylene glycol) precipitation procedure as described by the manufacturer (Terminator Cycle Sequencing Manual; Applied Biosystems, Inc.). Sequencing was performed by the chain termination method of Sanger et al. [13] using Sequenase 2.0 as recommended by the suppliers (U.S. Biochemical Corp.) in conjunction with both universal and specific primers.

Genomic library screening and plasmid construction

The 400 cosmid genomic library of *T. pantotropha* was that described by Berks et al. [14]. *Eco*RI fragments were separated by electrophoresis on a 0.5% agarose gel and denatured by 0.4 M NaOH solution. Single-stranded DNA fragments were transferred on to a piece of nylon membrane by Southern blotting [15] to permit probing by a digoxigenin-labelled PCR product corresponding to part of the pseudoazurin gene. The temperatures for hybridization and washing were 68°C and room temperature respectively.

A plasmid identified as containing the pseudoazurin gene was digested with *Eco*RI. The 4 kbp DNA fragment that included *pazS* was isolated from a 0.8% agarose gel [15] and was ligated into the *Eco*RI site of pUC18 [16]. The resulting plasmid was designated pYCL38.

The sequenced region of plasmid pYCL38 (Figure 1) showed convenient restriction sites for *Rsa*I [49 bp upstream of the ribosome-binding site (RBS)] and *Msc*I (95 bp downstream of the stop codon) on *pazS*. Thus digestion of pYCL38 with these restriction enzymes released a 0.6 kbp blunt-ended DNA fragment containing *pazS*, including the upstream regulatory sequences (see the Results section). This fragment was isolated from a 0.8% agarose gel as described above and ligated into the *Sma*I site of plasmid pUC18. The construct with the pseudoazurin gene in the same orientation as the *lacZ'* gene, and thus under the control of the *lac* promoter, was identified by restriction analysis

using *Eco*RI and *Nco*I double digestion, and is designated pYCL39. The plasmid with the pseudoazurin gene in the opposite orientation was designated pYCL40. The plasmids were transformed into *E. coli* strain XL1-Blue to study the level of expression of pseudoazurin.

Removal of the inverted repeat sequence upstream of the pseudoazurin gene

The 0.6 kbp DNA fragment containing *pazS* was excised from plasmid pYCL39 using the restriction enzymes *Xba*I and *Eco*RI and subcloned into the phagemid pBluescript KS[−] through sticky-end ligation to form plasmid pJR1. The latter was used for removing the 'hair-pin' sequence (see the Results section) by site-directed deletion mutagenesis using the uracil-template method [12] with a 20-base oligonucleotide, PAPp (5'-CCAG-GCACGAGACCACCAAG-3').

Twelve colonies were screened by DNA sequencing of the single-stranded DNA produced by infection with M13KO7 helper phage [17,18]. A recombinant plasmid selected after the screening was termed pJR2 and was used to transform *E. coli* strain XL1-Blue. Double-stranded sequencing employing the same methods as described above was used to check the gene for unwanted mutations.

Small-scale preparation of extracellular, periplasmic and cytoplasmic fractions

The *E. coli* strains used were XL1-Blue (pYCL39) and XL1-Blue (pYCL40). Cells from frozen stocks were inoculated into 2 ml of growth medium ($2 \times$ TY containing 0.1 mM CuSO_4 and $100 \mu\text{g/ml}$ ampicillin) and incubated at 37°C with shaking (200 rev./min) overnight. A 0.5 ml portion of the overnight culture was inoculated into 30 ml of the same growth medium in a 250 ml conical flask. When the A_{600} was equal to 1.0, isopropyl β -D-thiogalactoside (IPTG) was added as inducer to a final concentration of 0.3 mM. The cells were then grown for another 3 h before harvesting by centrifugation at 3800 g for 10 min at 4°C . The cell pellets were washed twice with 5 ml of cold 10 mM Tris/HCl buffer, pH 7.5, containing 25% (w/v) sucrose. The washed cells were then resuspended in 5 ml of the same buffer containing 25% (w/v) sucrose and 1 mM EDTA. After shaking the suspension for 10 min at room temperature, the cells were pelleted by centrifugation at 3800 g for 20 min at 4°C . The initial culture supernatant, the two washes and the supernatant after EDTA treatment constituted the extracellular protein fraction. The cells were then resuspended quickly and vigorously in 4 ml of ice-cold water. The suspension was shaken for 10 min at 4°C before being centrifuged at 3800 g for 20 min at 4°C . The supernatant contained all the soluble periplasmic proteins (periplasmic fraction). The pelleted cells were resuspended in 4 ml of ice-cold 10 mM Tris/HCl buffer, pH 7.5, and then disrupted by sonication ($3 \times 10 \text{ s}$; 170 W acoustic power at 20 kHz). Following centrifugation at 3800 g for 15 min at 4°C , the supernatant (cytoplasmic fraction) was collected.

Expression, extraction and purification of recombinant pseudoazurin from *E. coli* strains XL1-Blue (pYCL39) and XL1-Blue (pJR2)

An overnight culture (8 ml) of strain XL1-Blue (pYCL39) or XL1-Blue (pJR2) was inoculated into 400 ml of $2 \times$ TY medium containing $100 \mu\text{g/ml}$ ampicillin and 0.1 mM CuSO_4 in a 2-litre conical flask. The culture was grown at 37°C with shaking (200 rev./min). When the A_{600} was 1.0–1.2, expression was induced with 0.3 mM IPTG. Growth was continued for another 3 h

Figure 1 Restriction sites and sequencing strategy for the pseudoazurin gene and its flanking sequences in the 4 kbp *EcoRI* fragment from plasmid pYCL38

Small arrows indicate sequenced fragments using the primers listed in Figure 2. The shaded area shows the open reading frame.

before the cells (approx. 5 g wet weight per litre) were harvested by centrifugation at 3800 *g* for 15 min at 4 °C. The pelleted cells were frozen in solid CO₂ for 10 min and then thawed at room temperature for 20 min. This freezing and thawing cycle was repeated twice. Ice-cold 100 mM Tris/HCl buffer, pH 8 (30 ml), was added and the cell pellet was resuspended by shaking at 200 rev./min for 15 min at room temperature. The suspension was then centrifuged at 3800 *g* for 20 min at 4 °C and the supernatant (containing pseudoazurin) was separated from the pellet. The method resulted in the quantitative release of pseudoazurin from the periplasm.

The purification procedure for pseudoazurin was developed from that used in [9]. The proteins extracted from the periplasm were loaded on to a DEAE-Sephacrose CL6B ion-exchange column (1.6 cm × 18 cm) which had been equilibrated with 100 mM Tris/HCl buffer, pH 8, at 4 °C. After loading, the column was washed with 100 ml of the same buffer. Protein was eluted by a 0–300 mM NaCl linear gradient in the above buffer (total volume 250 ml). Fractions containing pseudoazurin were identified by their blue colour and by SDS/PAGE (10% gel), pooled and concentrated to a small volume (1–2 ml) using an Amicon concentrator with a YM10 ultrafiltration membrane. The concentrated sample was then loaded on to a Sephacryl S-200 gel-filtration column (1.6 cm × 100 cm) equilibrated with the same buffer, but containing 50 mM NaCl. Again, the blue colour and SDS/PAGE identified fractions containing pseudoazurin. Solid ammonium sulphate was added with stirring to the pooled fractions to 80% saturation. The precipitate was removed by centrifugation for 10 min at 3800 *g* at 4 °C and the blue pseudoazurin-containing supernatant was loaded on to a phenyl-Sepharose CL4B column (1.0 cm × 20 cm) that had been equilibrated with 80% saturated ammonium sulphate in 100 mM Tris/HCl buffer, pH 8, at 4 °C. The column was eluted with a linear gradient of 80–0% saturated ammonium sulphate in the same buffer. Pseudoazurin was eluted at 30% saturated ammonium sulphate. Fractions containing pseudoazurin identified as above were pooled. The sample was then exchanged into 100 mM Tris/HCl buffer, pH 8, by ultrafiltration and stored at –20 °C. For long-term storage, the protein was freeze-dried and the powder stored at –20 °C. The purified recombinant pseudoazurin gave a single band on SDS/PAGE of the same mobility as the wild-type protein isolated from *T. pantotrophus*. The protein concentration was calculated using $\epsilon_{590} = 1.36 \text{ mM}^{-1} \cdot \text{cm}^{-1}$ [4].

N-terminal amino acid sequence analysis

The first 10 amino acids from the N-terminus of pseudoazurin (10 µg) purified from strains XL1-Blue (pYCL39) and XL1-Blue

(pJR2) were determined by stepwise Edman degradation using a gas-phase sequencer (Applied Biosystems; model 470A).

Purification and sequencing of pseudoazurin from *Paracoccus denitrificans* 8944

The protein was purified and fragmented for direct sequence analysis essentially as described previously for pseudoazurin from *P. pantotrophus* [9]. The sequence has been deposited under accession no. P80649 in the EMBL Protein Sequence Database.

Measurement of electrode potential

The electrode potentials of the recombinant and wild-type pseudoazurins were measured by direct electrochemistry [19], as described by Moir et al. [4].

Analysis of transcription of *pazS* by primer extension

Total RNA from *T. pantotrophus* grown under denitrifying conditions [9] was extracted by the hot phenol method described by Aiba et al. [20]. An antisense primer of 30 nucleotides with the sequence 5'-GGGCGCGGCAAGGGCCAGAAGCGCGGC-GGC-3', designed based on the DNA sequence approx. 50 bp downstream of the RBS, was radiolabelled with [γ -³²P]ATP using polynucleotide kinase (New England Biolabs) according to the manufacturer's instructions. The primer was then hybridized with the template RNA and subjected to reverse transcription with Moloney murine leukaemia virus reverse transcriptase (Promega), essentially as described by Sambrook et al. [11]. The primer extension product was then run alongside a DNA sequence ladder generated with the same primer.

RESULTS AND DISCUSSION

Cloning and sequencing of the pseudoazurin gene

At the outset of this work, a partial amino acid sequence of pseudoazurin from *T. pantotrophus* had been determined. Four highly degenerate oligonucleotides were synthesized on the basis of the known amino acid sequence of pseudoazurin from *T. pantotrophus*, two of which successfully acted as primers for the PCR. The combination of PAPa [5'-GT(A/C/G/T)CA(C/T)-ATG(C/T)T(A/C/G/T)AA(C/T)AA(A/G)GG(A/C/G/T)GA-3'] as forward primer, based on the N-terminal sequence (VHMLNKG), and PAPb [3'-GG(A/C/G/T)GT(A/G)AA(A/G)CC(A/C/G/T)TACCC(A/C/G/T)TACCA-5'] as reverse primer, based on a region close to the C-terminal sequence (PHFGMG MV), generated a PCR product of the expected size (approx. 250 bp), which was subcloned into pBluescript KS[–] to give plasmid pTL3. DNA sequencing confirmed that the fragment corresponded to the predicted part of the pseudoazurin gene. Digestion of pTL3 with *PvuII* released the PCR product, which was labelled with digoxigenin following purification with GeneClean. This digoxigenin-labelled probe was used to identify a 4 kb *EcoRI* fragment of a cosmid that contained the pseudoazurin gene. The fragment was subcloned into plasmid pUC18 to generate pYCL38. A series of oligonucleotides (Figure 2) was used to sequence the entire gene and its flanking region on both strands of DNA (Figure 1). The 1084 bp sequenced is G + C-rich (62.9%) (Figure 3), as reported for most genes cloned from the α -3 group of bacteria that includes *T. pantotrophus* [21]. The relatively high GC content caused occasional difficulties in

PAPc	5' - CCATGGCGCCCGATTGCCC - 3'
	470 451
PAPd	5' - GAAATGCACGCCGCACTTCG - 3'
	649 668
PAPe	5' - GCGACATTCTGTTCATGCC - 3'
	223 240
PAPf	5' - GGCATGACAGAATGTCGC - 3'
	240 223
PAPg	5' - CGACCGTCCGGGAGAGGG - 3'
	806 822
PAPi	5' - CCCTCTCCCGGACGGTCG - 3'
	822 806
PAPj	5' - CTCCGGTCTCGCAGCG - 3'
	1083 1068
PAPk	5' - GCCGCGCCCGGTTTCGCG - 3'
	404 421
PAPl	5' - GCCGAAATGCGGCGTGC - 3'
	670 654
PAPm	5' - GTGCCGGCCCATCCCGCC - 3'
	29 46
PAPn	5' - GACAAAGGCATAATGAGAGG - 3'
	140 121
PAPo	5' - GTCGATCTTGCTCATGCC - 3'
	964 947

Figure 2 Oligonucleotide primers used for sequencing of the pseudoazurin gene and its flanking regions

The primers (e.g. PAPg, which has an extra base) do not necessarily correspond exactly to the final sequence.

reading sequences, but these could always be unambiguously resolved by sequencing the complementary strand. The primers listed in Figure 2 were generally designed on the basis of sequence from one strand alone as part of an efficient primer walking approach. Of these, PAPd and PAPg proved not to match the template DNA faithfully. Nevertheless, they both bound very well to the template and generated good-quality sequence data for the adjacent DNA; the errors in PAPd and PAPg became clear from sequencing the complementary strand.

The sequenced fragment contains only one open reading frame, with a Shine–Dalgarno consensus sequence or RBS (boxed in Figure 3) at the expected distance (5 nucleotides) from the start codon ATG. At 9 bp upstream of the RBS, there is a possible inverted repeat DNA sequence or ‘hairpin’-forming region. A comparison of the open reading frame with the sequence of the mature pseudoazurin protein [9] showed no differences, except for the presence of a region in the gene encoding a signal

Figure 3 DNA sequence of the pseudoazurin gene and deduced amino acid sequence of the protein

The nucleotides are numbered. Possible promoter consensus sequences (*ntf A* and *fmr/nmr*) are boxed and labelled. The termination codon is indicated by an asterisk. Hairpin-forming regions are indicated by arrows above the DNA sequence. The RBS or Shine–Dalgarno consensus sequence is boxed and labelled (SD). The processing site of the precursor protein is indicated. Restriction sites for *SmaI*, *RsaI* and *MscI* are underlined and labelled.

sequence of 22 amino acids (Figure 3). This region exhibits all the characteristics of a prokaryotic signal peptide [22]. It contains two charged residues (two His residues at positions –19 and –20) towards the N-terminus, a long hydrophobic central region (positions –5 to –17), and two β -turn-promoting residues (Pro

at -4 and Gly at -3) towards the C-terminus of the propeptide. Alanine residues typical of the small amino acids observed at the processing site are also seen in this region (Ala at positions -1 and +1). The central hydrophobic region of the signal sequence is unusual, however, in that it contains only two amino acid types, alanine (nine residues) and leucine (four residues). In studies of the consequences of mutating the hydrophobic region of the signal sequence for *E. coli* alkaline phosphatase, it has been shown that, although a sequence of 10 leucine residues will effectively direct the enzyme to the periplasm, progressive introduction of alanines causes an attenuation of translocation [23]. While the attenuation was of the order of 20% at an Ala/Leu ratio of 1:1, a further incremental increase in the alanine content to a ratio of 3:2 caused, in general, a decrease to 30% of the maximum translocation efficiency. This result was obtained with the two sequences LAALLAALAA and LALAAAALAL and, as further incorporation of Ala resulted in additional progressive loss of translocation activity, it was argued that there is a critical requirement for hydrophobicity as reflected in the alanine/leucine ratio [23]. However, Doud et al. [23] also noted that the sequence LALAAALALA was competent at supporting 80% of normal translocation activity, suggesting that hydrophobicity, although very important, does not necessarily directly correlate with translocation activity. The signal sequence, active in *E. coli* (see later), identified here for *T. pantotrophus* pseudoazurin possesses alanine in a greater ratio than 2:1 over leucine. While the pseudoazurin hydrophobic core is 13 residues long, compared with 10 in alkaline phosphatase, and a longer hydrophobic core in the signal peptide does increase the translocation activity [24], the present findings do indicate that there may not be a straightforward general relationship between hydrophobicity and translocation, although we cannot rule out the possibility that pseudoazurin from *T. pantotrophus*, *Alcaligenes faecalis* S-6 [25] and *Achromobacter cycloclastes* [26] (where similar signal sequences can be deduced) has adopted a less than maximally active signal sequence.

There are two other ATG start codons, at positions 116–118 and 236–238, which are in the same reading frame as the ATG identified in Figure 3. These two alternative translational initiation codons are discounted because they would mean an unusually long and atypical periplasmic target sequence, they are not adjacent to an RBS and, finally, they are inconsistent with the transcription start point that is identified later.

An inverted repeat sequence was found 28 bp downstream of the stop codon (TGA) at the end of the gene. This is probably a transcriptional terminator, which may act as a transcriptional pause signal [27].

Among denitrifying organisms that contain pseudoazurin, *T. pantotrophus* is clearly closely related to *Paracoccus denitrificans*, but there is currently debate as to whether these two organisms are distinct species ([7]; I. P. Thompson, S. J. Ferguson and S. C. Baker, unpublished work). For this reason it was of interest to acquire the primary sequence of *P. denitrificans* pseudoazurin. This was done by direct protein sequencing, with the result that just six changes from the *T. pantotrophus* sequence, obtained here from the gene and presented previously from protein sequencing [9], were identified. The changes are V23I, V34I, S55T, S63A, T65A and K103Q, and support the similarity rather than identity of the two organisms.

Upstream regulatory regions in the pseudoazurin gene

Information about a possible promoter and regulatory sequences upstream of *pazS* was sought via identification of the transcription start point through a primer extension experiment

Figure 4 Primer extension analysis for the *pazS* gene

The same labelled oligonucleotide as that used for primer extension was used to generate the sequencing ladders in lanes T, G, C and A. The sequence shown is that of the coding strand. The base representing the start of transcription in lane 1 is indicated by an arrow. This gel was not used for acquisition of primary DNA sequence data. Sequencing gels gave higher quality data for the region shown here and, as explained in the text, any ambiguities were readily resolved by sequencing the complementary strand.

using reverse transcriptase. The one major product observed (Figure 4) indicates that transcription is initiated at the guanosine that is 31 nucleotides upstream of the RBS. Characteristic promoter motifs can frequently be seen centred at 10 and 35 bases upstream of the transcription initiation point for bacterial genes. As mentioned above, *T. pantotrophus* is closely related to *P. denitrificans* and other organisms of the α -3 subclass. On the basis of sequence analysis of relatively few genes a consensus -10, -35 box, TCGGGNN_(18±2)GATNG(C/G), has been proposed for these organisms [21]. A recent study in our laboratory has identified such a putative promoter for the cytochrome *c*₅₅₀ gene of *P. denitrificans* [28]. Inspection of the DNA sequence upstream of the transcription start point for pseudoazurin shows that the provisional consensus -10, -35 sequence for an organism such as *T. pantotrophus* is absent. There is also no resemblance in this region to the typical consensus sequence for the *E. coli* σ ⁷⁰ factor.

The most recognizable DNA sequence upstream (centred at position -36) of the transcription start site in the pseudoazurin gene is CTTTGACTTTTCTCAAGC. This is clearly similar to the consensus *fur* box sequence, TTGATN₄ATCAA, that has been identified in *E. coli* [29]. It is known that the expression of pseudoazurin in *T. pantotrophus* is strictly dependent on anaerobic growth conditions [10], and thus a sequence analogous to

Figure 5 Analysis by SDS/PAGE of the distribution among cell fractions of pseudoazurin overexpressed in *E. coli*

Lane 1 contains 2 μ g of wild-type pseudoazurin from *T. pantotropha*. The material in lanes 2–4 was obtained from a culture of cells expressing plasmid pYCL39. The loading of each lane was normalized such that equivalent proportions of the different cell fractions (lane 2, extracellular; lane 3, periplasm; lane 4, cytoplasm) were analysed. The material in lanes 5 (extracellular), 6 (periplasm) and 7 (cytoplasm) was obtained similarly from a culture of cells expressing plasmid pYCL40. Lane 8 contains marker proteins (1 μ g of each protein): lysozyme (14 300 Da), trypsin inhibitor (21 500 Da), carbonic anhydrase (31 000 Da), ovalbumin (45 000 Da), BSA (66 200 Da) and phosphorylase *b* (97 400 Da).

the *fnr* box (to which, in *E. coli*, the Fnr protein binds and thereby regulates transcription of a set of genes under anaerobic conditions) is not unexpected. A comparable sequence is seen upstream of another anaerobically expressed gene in *T. pantotropha*, i.e. *nirS*, coding for nitrite reductase (N. F. W. Saunders, S. C. Baker and S. J. Ferguson, unpublished work). It has been argued that two closely related proteins in *P. denitrificans*, Fnr and Nnr [30], are involved in regulating anaerobic gene expression. It is very probable that one, or even both, of these proteins is involved in regulating the expression of the pseudoazurin gene in *T. pantotropha*. This may be a widespread feature of azurin and pseudoazurin gene expression, because comparable *fnr*-type recognition sequences have been identified in the upstream region of the *Alcaligenes denitrificans* azurin gene (GATTGATGTCCTCAATA), the *Pseudomonas aeruginosa* azurin gene (GTTTGACCTGAATCAGTG) and the *Alcaligenes faecalis* pseudoazurin gene (GTTTGATCCAGATCAAAG).

All of the above considerations point to a role for the *fnr*/*nnr* box in regulating transcription of the pseudoazurin gene. In *E. coli* the *fnr* box is usually found between 3 and 13 bases upstream of the -35 box that is recognized by typical σ^{70} -dependent promoters. As noted earlier, there is no candidate promoter of either the *E. coli* or α -3 subgroup of bacteria types in the expected position for the *T. pantotropha* pseudoazurin gene. This raises two possibilities. The first is that there could be a novel type of promoter recognition site somewhere between the *fnr*/*nnr* box and the transcription start point. The second is that the promoter recognition site is elsewhere. In this regard, it is notable that a sequence similar to the consensus *ntrA* box described for enteric bacteria can be identified in the upstream region of not only the *T. pantotropha* *pazS* gene (Figure 3) but also the azurin and pseudoazurin genes of *Alcaligenes denitrificans*, *Pseudomonas aeruginosa* and *Alcaligenes faecalis* [25,31,32]. In enteric bacteria an *ntrA* box is usually found at -24 and -11 bases upstream of the transcription start point, and serves as a recognition site for an alternative sigma factor, σ^{54} . If the putative *ntrA* box identified upstream of the pseudoazurin and azurin genes serves a similar role, then its relationship

with the transcription start point is clearly quite novel. In this context it is nevertheless worth noting that, in enteric bacteria, the activity of the *ntrA*-dependent transcription process is regulated by DNA sequences many tens of bases upstream, as is the case for eukaryotic enhancers, which implies that bending of DNA is very significant. While the finding of a putative *ntrA* site far upstream of the transcription start point for azurins and pseudoazurin may be coincidental, it is possible that this is a novel example of a relatively distal promoter binding site which, through bending of DNA, is able to direct an RNA polymerase to the correct transcription start point.

Overexpression of recombinant pseudoazurin in *E. coli*

When cells containing pYCL39 were induced with IPTG, more than 90% of the total pseudoazurin produced was located in the periplasm (Figure 5). The remaining material was found in the extracellular medium. By contrast, when the pseudoazurin gene was present in the opposite orientation (plasmid pYCL40), no pseudoazurin was seen in the extracellular, periplasmic or cytoplasmic fractions of the cells (Figure 5). In *T. pantotropha*, the original Gram-negative bacterial host for pseudoazurin, the protein functions as an electron transporter in the periplasm [4]. The translocation apparatus of *E. coli* is evidently able to recognize the N-terminal signal sequence of the *T. pantotropha* protein and directs the protein to the periplasm. Using these conditions, after purification approx. 40 mg of pseudoazurin/litre of culture was obtained.

Examination of the region commencing 19 bp upstream from the start codon of the pseudoazurin gene revealed a palindromic sequence (-20 to -39) (Figure 3), predicted to correlate with a putative hairpin structure in the transcribed mRNA. A similar hairpin sequence upstream of the *Alcaligenes faecalis* S-6 pseudoazurin gene has been identified, the removal of which improves the level of expression of the protein in *E. coli* [25].

In an attempt to improve the yield of pseudoazurin from *T. pantotropha* in an expression system, therefore, the region containing the putative hairpin was also deleted using the oligonucleotide PAPp. The hairpin region was removed by deletion mutagenesis. Twelve potential mutants were screened, of which six had the correct sequence. Double-stranded sequencing of the resulting gene introduced into pBluescript KS[−] (pJR2) showed that there were no additional, aberrant, mutations in the gene. Subsequent expression of the protein from this plasmid resulted in an expression level approximately twice that obtained with the original pseudoazurin gene (80 mg of purified protein/litre). Over 90% of the pseudoazurin was still targeted to the periplasm.

Characterization of recombinant pseudoazurin

The purified recombinant pseudoazurin had the same mobility on SDS/PAGE as the wild-type protein purified from *T. pantotropha*. The N-terminal sequence was Ala-Thr-His-Glu-Val-His-Met-Leu-Asn-Lys-, identical with that of pseudoazurin from *T. pantotropha*. This demonstrated not only that the protein had been transported to the periplasm, but also that the signal sequence had been processed correctly.

Structural and functional analysis of recombinant pseudoazurin also showed that it was indistinguishable from its wild-type counterpart. The standard electrode potentials of both the wild-type and recombinant proteins were determined by a non-mediated system using an edge-planed graphite electrode [19], and both gave the same value of 270 mV. A functional assay using pseudoazurin as electron donor to a nitrite reductase also indicated that it is redox-active (C. Chan and S. J. Ferguson,

unpublished work). Similarly, visible absorption spectra of the purified recombinant protein showed the characteristic peaks expected for pseudoazurin at 450 nm, 590 nm and 760 nm [4]. The far- and near-UV CD spectra and EPR spectra were also essentially identical with those of the wild-type protein (results not shown). Furthermore, when the recombinant protein was analysed by electrospray ionization MS, only one species was present, the mass of which (13342 ± 1 Da) was identical with that of pseudoazurin purified from *T. pantotropha*. Electrospray ionization MS under native conditions confirmed the presence of a single bound copper ion (13402 Da). Interestingly, species containing an oxidized methionine, found in preparations of the protein obtained from *T. pantotropha* [9], were not observed in the recombinant protein.

Taken together, therefore, the present data indicate that recombinant pseudoazurin has the same fold and activity as the wild-type protein. The enhanced level of expression when *E. coli* is used as a host, in combination with an improved purification procedure, sets the stage for a detailed analysis of the conformational changes occurring upon reduction of the copper site, the role of pseudoazurin in electron transport to nitrite reductase of *T. pantotropha*, the role of specific amino acids in this molecular recognition event and the structural determinants of the Greek-key motif. Furthermore, the nucleotide sequence of the pseudoazurin gene and of its promoter regions now makes it possible to study the regulation of the expression of the pseudoazurin gene in detail.

We thank Dr. David Pearce for his advice concerning the cloning of pseudoazurin. This is a contribution from the Oxford Centre for Molecular Sciences, which is supported by BBSRC, EPSRC and MRC. We thank the BBSRC for studentships (C.C. and J.S.R.), the Royal Society for supporting S.E.R. and the Croucher Foundation for a fellowship to Y-C.L. We are grateful to Carol Robinson for making the measurements by electrospray ionization MS.

REFERENCES

- Berks, B. C., Ferguson, S. J., Moir, J. W. B. and Richardson, D. J. (1995) *Biochim. Biophys. Acta* **1261**, 279–284
- Adman, E. T. (1991) *Adv. Protein Chem.* **42**, 145–197
- Berks, B. C., Baratta, D., Richardson, D. J. and Ferguson, S. J. (1993) *Eur. J. Biochem.* **212**, 467–476
- Moir, J. W. B., Baratta, D., Richardson, D. J. and Ferguson, S. J. (1993) *Eur. J. Biochem.* **212**, 377–385
- Moir, J. W. B. and Ferguson, S. J. (1994) *Microbiology* **140**, 389–397
- Williams, P., Fulop, V., Leung, Y. C., Chan, C., Moir, J. W. B., Howlett, G., Ferguson, S. J., Radford, S. E. and Hajdu, J. (1995) *Nature Struct. Biol.* **2**, 975–982
- Goodhew, C. F., Pettigrew, G. W., Devreese, B., van Beeumen, J., van Spanning, R. J. M., Baker, S. C., Saunders, N., Ferguson, S. J. and Thompson, I. P. (1996) *FEMS Microbiol. Lett.* **137**, 95–101
- Reference deleted
- Chan, C., Willis, A. C., Robinson, C. V., Aplin, R. T., Radford, S. E. and Ferguson, S. J. (1995) *Biochem. J.* **308**, 585–590
- Moir, J. W. B., Richardson, D. J. and Ferguson, S. J. (1995) *Arch. Microbiol.* **164**, 43–49
- Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd edn., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
- Kunkel, T. A., Roberts, J. D. and Zakour, R. A. (1987) *Methods Enzymol.* **154**, 367–382
- Sanger, F., Nicklen, S. and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5463–5467
- Berks, B. C., Richardson, D. J., Reilly, A., Willis, A. C. and Ferguson, S. J. (1995) *Biochem. J.* **309**, 983–992
- Southern, E. M. (1975) *J. Mol. Biol.* **98**, 503–517
- Errington, J. (1990) *Nucleic Acids Res.* **18**, 5324
- Messing, J. (1983) *Methods Enzymol.* **101**, 20–79
- Vieira, J. and Messing, J. (1987) *Methods Enzymol.* **153**, 3–11
- Frew, J. E. and Hill, H. A. O. (1988) *Eur. J. Biochem.* **172**, 261–269
- Aiba, H., Adhya, S. and de Crombrugghes, B. (1981) *J. Biol. Chem.* **256**, 11905–11910
- Steinrück, P. and Ludwig, B. (1993) *FEMS Microbiol. Rev.* **104**, 83–118
- Duffaud, G. D., Lehnhardt, S. K., March, P. E. and Inouye, M. (1985) *Curr. Top. Membr. Transp.* **24**, 65–104
- Doud, S. K., Chou, M. M. and Kendall, D. A. (1993) *Biochemistry* **32**, 1251–1256
- Chou, M. M. and Kendall, D. A. (1990) *J. Biol. Chem.* **265**, 2873–2880
- Yamamoto, K., Uozumi, T. and Beppu, T. (1987) *J. Bacteriol.* **169**, 5648–5652
- Chen, J.-Y., Chang, W.-C., Chang, T., Chang, W.-C., Liu, M.-Y., Payne, W. J. and LeGall, J. (1996) *Biochem. Biophys. Res. Commun.* **219**, 423–428
- Silvestrini, M. C., Galeotti, C. L., Gervais, M., Schinina, E., Barra, D., Bossa, F. and Brunori, M. (1989) *FEBS Lett.* **254**, 33–38
- Stoll, R., Page, M. L. D., Sambongi, Y. and Ferguson, S. J. (1996) *Microbiology* **142**, 2577–2585
- Spiro, S. and Guest, J. R. (1987) *J. Gen. Microbiol.* **133**, 3279–3288
- van Spanning, R. J. M., de Boer, A. P. N., Reijnders, W. N. M., Spiro, S., Westerhoff, H. V., Stouthamer, A. H. and van der Oost, J. (1995) *FEBS Lett.* **360**, 151–154
- Hoitink, C. W. G., Woudt, L. P., Turenhout, J. C. M., van de Kamp, M. and Canters, G. W. (1990) *Gene* **90**, 15–20
- Arvidsson, R. H. A., Nordling, M. and Lundberg, L. G. (1989) *Eur. J. Biochem.* **179**, 195–200