The structural gene for rusticyanin from *Thiobacillus ferrooxidans*: cloning and sequencing of the rusticyanin gene

John F. Hall \(^a\), S. Samar Hasnain \(^{a,b,}\)*, W. John Ingledew \(^c\)

\(^a\) School of Applied Sciences, De Montfort University, Leicester LE1 9BH, UK
\(^b\) Molecular Biophysics Group, Daresbury Research Laboratory, Warrington, UK
\(^c\) School of Biological and Medical Sciences, University of St. Andrews, St. Andrews KY16 9AL, UK

Received 12 October 1995; revised 19 January 1996; accepted 19 January 1996

**Abstract**

The rusticyanin gene from the acidophilic chemolithotroph *Thiobacillus ferrooxidans* has been cloned and sequenced. A central portion of the gene was identified by PCR reactions utilising primers optimised for codon bias followed by nested PCR with degenerate primers. The 5' and 3' ends of the rusticyanin gene were then cloned using degenerate primers to each end and anchor sequences to the known internal sequence. The entire gene was amplified using *Tli* DNA polymerase and specific primers to the 5' and 3' ends and the sequence confirmed after cloning into Bluescript and transformation of XL-1 Blue *Escherichia coli*.

**Keywords:** Rusticyanin; Gene cloning; *Thiobacillus ferrooxidans*; Redox protein

**1. Introduction**

Rusticyanin is a small Type-I copper protein and is thought to be a principal component in the iron respiratory electron transport chain of *Thiobacillus ferrooxidans* [1,2]. The protein was first identified and partially characterised by Cobley and Haddock [3] from *T. ferrooxidans* grown on the oxidation of ferrous iron, expression of rusticyanin is dependent on this growth phenotype [4,5]. Rusticyanin is a member of a group of proteins called small (12-17 kDa) Type-I ‘blue copper’ proteins [6]. These proteins are associated with electron transfer systems and have been extensively studied as model electron transfer proteins and contain a single copper as the prosthetic group which endows these proteins with a blue colour. The copper centre undergoes oxidation and reduction between the Cu(II) and Cu(I) states with rapid turnover, thus invoking the idea of entatic state in which the Cu site is strained to non-ideal geometry for the Cu(II) condition. Within this group, rusticyanin has the unusual feature of extreme acid stability and possesses the highest redox potential within the family. Thus it is of considerable interest to elucidate its detailed atomic structure.

The crystallographically characterised blue proteins have shown a basic framework of a copper site in which copper ligation is provided by 2 His, 1 Cys and 1 Met. The lack of methionine in Stellacyanin is...
thought to be responsible for the lowest redox potential, 184 mV, in the family and has been extensively investigated by site-directed mutagenesis [7–9] of azurins from *Pseudomonas aeruginosa* and *Alcaligenes denitrificans*. The amino acid sequence of rusticyanin has a very low invariance (< 20%) from the mesophiles of its class. Amino acid sequences of rusticyanin from four strains of *T. ferrooxidans* are known and show a high degree of homology (> 90%) among themselves [2,10–12]. These sequence data taken in combination with the results of physical studies have enabled identification of three of the ligands to the copper prosthetic group located at the C-terminus. There is some conjecture over the nature of the putative fourth ligand and several alternatives (His 39, 57 or 85, Met 148 and Asp 73) have been suggested. A recent analysis [13], combining EXAFS studies and homology modelling, has suggested the fourth ligand to be histidine, specifically His 85. These authors have also suggested that the substitution of serine for asparagine (which is conserved in other type-1 "blue copper proteins") next to His 85 accounts, in part, for the high stability of the copper site. A useful approach to test this model and clarify further the structural determinants which impart the extreme properties on rusticyanin is to study the effects of amino acid substitutions in the protein. A prerequisite for this is the isolation and sequencing of the gene encoding rusticyanin, which is reported here.

2. Materials and methods

2.1. Cell growth and maintenance

*T. ferrooxidans* cells were grown on a ferrous iron source in continuous chemostat culture as previously described [1]. The cultures were maintained in shake flasks at 25°C in a medium comprised of solutions 1 and 2 mixed in a ratio of 4:1. Solution 1 was autoclaved and consisted of: K$_2$HPO$_4$, 0.5 g; (NH$_4$)$_2$SO$_4$, 0.5 g; MgSO$_4$, 0.5 g; 0.5 M H$_2$SO$_4$, 5.0 ml made to 1 l with deionised water. Solution 2 was filter-sterilised and consisted of FeSO$_4$·7H$_2$O, 167 g; 0.5 M H$_2$SO$_4$, 50 ml made to 1 l with deionised water.

pBluescript II KS ± phagemids were maintained in XL-1 Blue *Escherichia coli* (Stratagene) grown in LB medium containing ampicillin (50 μg ml$^{-1}$) and tetracycline (20 μg ml$^{-1}$).

2.2. Preparation of DNA

The *T. ferrooxidans* cell pellet (approx. 1.0 g wet weight) was washed sequentially by resuspension and re-centrifugation in solution 1 (50 ml), 50 mM pH 8.0 Tris buffer (50 ml) and 0.1 M pH 8.0 EDTA (50 ml). DNA was then prepared using the following adaptation of the standard method normally used for plasmid isolation [9][14] and K.M.A. Gartland and J. Gartland, personal communication. (1) The pellet was resuspended in 12.5 ml of 25% sucrose in 50 mM Tris pH 8.0. (2) 4 ml of 10 mg ml$^{-1}$ lysozyme in 50 mM Tris pH 8.0 were added and incubated for 5–10 min on ice after mixing. (3) 4 ml of 0.5 M EDTA pH 8.0 were added, mixed thoroughly and incubated on ice for a further 5–10 min. (4) 20 ml of 2% Triton X-100 in 50 mM Tris pH 8.0 were added and mixed vigorously, equilibrated for 20 min and remixed. (5) The suspension was then centrifuged for 60 min at 39 000 × g at 4°C. (6) The supernatant (30 ml) was carefully removed using a 5-ml Gilson pipette to a fresh 50-ml centrifuge tube to which 30 g of CsCl$_2$ and 0.75 ml of 10 mg ml$^{-1}$ ethidium bromide were added. The CsCl$_2$ was dissolved by shaking gently and the solution left in the dark for 30 min at room temperature. (7) Following centrifugation at 12000 × g for 30 min at 20°C the liquid fraction was transferred to an ultracentrifuge tube very carefully with a Pasteur pipette from inside a 5-ml Gilson tip which was inserted into the tube after pushing the thick layer on top to one side. This is relatively easy with *T. ferrooxidans* preparations as the crust is almost solid. (8) This liquid fraction was then centrifuged at 150000 × g for 16 h at 20°C in a vertical rotor. (9) The darkly stained band visualised by UV illumination if necessary was removed and purified following standard procedures [14].

2.3. PCR

PCR was performed on either a COY thermocycler or an MJ Research minicycler. Oligonucleotides
were synthesised on an Applied Biosystems 391 DNA synthesizer PCR-Mate. Taq DNA polymerase was used routinely for PCR amplification and Tli DNA polymerase, with an integral 3'-5' exonuclease or 'proofreading' function used in the final amplification of the entire gene to minimize errors.

The conditions used for PCR were: Taq DNA polymerase (1:5 or 1:10 dilution) 1 μl, ×10 MgCl₂ 1 μl, ×10 buffer 2 μl, ×10 dNTPs 2 μl, target DNA 1 μl, primers were used at between 70 and 80 pmol and the reaction mixture made to a final volume of 20 μl with deionised water. The conditions were altered slightly for Tli DNA polymerase such that it was used diluted 1:5 and 4 μl of ×10 MgCl₂ were added. The final concentrations of MgCl₂ were therefore 1.25 mM and 5 mM for Taq and Tli, respectively. The final concentrations of the other components were: dNTPs 0.2 mM, buffer, 50 mM KCl, 10 mM Tris·HCl pH 9.0 at 25°C, 0.1% Triton X-100.

Target DNA from T. ferrooxidans was sheared by vortexing and resuspension through a 1-ml Gilson tip prior to use as a PCR template. This was an important step to obtain appropriate products from the PCR reactions.

Primer 1: AAGGACACCGGCAAGG
Primer 2: AGGTGAAATCGTGGTGAAGC
Primer 3: TTI(CG)GAG(AGC)GT(AGC)CTGAA
Primer 4: C(C)T(TT)AT(AGC)TGT(AG)AT(AG)TC
Primer 5: GGNACNNTT(AG)GACTACNTGGA(AG)GA(AG)GC
Primer 6: (CT)TNTACNAC(AG)AT(TT)TNCC(AG)AAATNCCNGT
Primer 7: TTCAAGGATTGACACAAAAAG
Primer 8: TTTTTTGATGATGTCAAAAACT
Primer 9: GTTACATTGGATACATCCAG
Primer 10: CTTAACCAAGATCTTTGGCAGA

Brackets indicate degeneracy at this position. N indicates complete degeneracy at this position.

2.4. Gel electrophoresis

PCR products were separated using 2.5–3.5% Nusieve GTG agarose (FMC Bioproducts-Flowgen) with HaeIII digested pAT 153 as markers. Bands identified by their size were purified using Geneclean II (BIO 101 Inc.) and polished simultaneously with Klenow and kinase following the protocol of Starr and Quaranta [15]. The PCR products were purified using Geneclean II for a second time and used in ligation reactions with Smal cut Bluescript. XL-1 Blue E. coli cells were made competent and transformed using a method slightly modified from that of Nishimura et al. [16].

2.5. DNA sequencing

Plasmids containing the PCR products to be sequenced were purified using Wizard minipreps (Promega). The Sequenase II kit (USB, Amersham) was used according to the manufacturers instructions with [³⁵S]dATP as the radiolabelled nucleotide. Synthetic oligonucleotides M13, T7 or T3 were used as the primers in combination with primers complementary to part of the original sequence (from amino acids 54–93) when necessary. Manganese buffer (USB, Amersham) was used to enable reading of the sequence close to the primer at the 3' end of the gene, this was essential to detect bp 421 which otherwise appeared to be missing. Compressions around the GC-rich Ala, Ala, Ala, region (amino acids 42–44) were overcome after sequencing in both directions.

2.6. Screening of PCR products and recombinant plasmids

Following selection of possible recombinant plasmids on LB plates containing 50 μg ml⁻¹ ampicillin, 20 μg ml⁻¹ tetracycline and IPTG/X-GAL (isopropyl-β-d-thiogalactopyranoside and 5-bromo-4-chloro-3-indoly-β-d-galactopyranoside) using the blue/white selection system in Bluescript, plasmids were prepared using alkaline lysis, restricted with EcoRI and BamHI after which RNA was removed with Rnase-it (Stratagene). These preparations were then subjected to gel electrophoresis as above.

DIG-labelled probes were prepared using the random or end labelling kit supplied by Boehringer Mannheim. Southern transfers were performed directly in 0.4 M NaOH onto positively charged nylon membranes (Hibond N + Amersham). Hybridisations were performed with minor alterations to the standard protocol and detected using the NBT/X-Phos system (Boehringer Mannheim).
3. Results and discussion

The primary difficulty encountered was in designing primers from the highly degenerate sequence obtained by reverse translation of the known amino acid sequence. Primers 1 and 2 were synthesised as 'best guess' sequences after consideration of codon usage by *T. ferrooxidans* [17]. PCR under low stringency conditions: denatured at 95°C for 30 s, annealed at 40°C for 2 min and extended at 65°C for 2 min for 30 cycles gave a large number of products several of which were of the anticipated size (approx. 318 bp). These were excised from the gel and purified as described above before use as the template for nested PCR using the degenerate primers 3 and 4 (internal to the original primer pair) under the same low stringency conditions but with a final extension step of 7 min at 65°C. This produced a few bands after gel electrophoresis, one of which corresponded in size to the expected product of approx. 118 bp.

This band was excised, purified, and cloned into the *SmaI* site of Bluescript. Sequence analysis showed that this product coded for the central region of the gene from amino acids 54–93 as expected. In order to ascertain the entire sequence completely degenerate primers (5 and 6) were synthesised to either end of the gene as predicted from the published amino acid sequence. These were used with sheared DNA in PCR reactions under the following conditions: 95°C, 30 s; 50°C, 60 s; 65°C, 60 s; 40 cycles with a final extension at 65°C for 5 min. This gave one major band on gel electrophoresis (approx. 460 bp) which contained the desired sequence as demonstrated by blotting and hybridisation with the DIG-labelled central sequence previously isolated and nested PCR with primers 7 and 8, but was insufficiently pure for direct cloning and sequencing. These products were purified and used as the template for anchored PCR utilising the primer pairs 5 (5' end, degenerate) and 8 (specific to the 3' end of the known internal sequence) and 6 (3' end, degenerate) and 7 (specific to the 5' end of the known internal sequence) to amplify the gene in two fragments. PCR with these primers gave single products of sizes 276 bp (primers 5 and 8: 95°C, 30 s; 50°C, 60 s; 65°C, 60 s; for 30 cycles) and 305 bp (primers 6 and 7: 95°C, 30 s; 55°C, 60 s; 65°C, 60 s; for 30 cycles) as anticipated.

Several clones containing these products were isolated of which six of each were sequenced as detailed above. The sequences of the two clones overlapped completely from primer 7 to 8 and gave the full sequence for the gene encoding rusticyanin

```
GGTACATTGGATACTACATGGAAAGAGGCGACGCTT
```

This band was excised, purified, and cloned into the *SmaI* site of Bluescript. Sequence analysis showed that this product coded for the central region of the gene from amino acids 54–93 as expected. In order to ascertain the entire sequence completely degenerate primers (5 and 6) were synthesised to either end of the gene as predicted from the published amino acid sequence. These were used with sheared DNA in PCR reactions under the following conditions: 95°C, 30 s; 50°C, 60 s; 65°C, 60 s; 40 cycles with a final extension at 65°C for 5 min. This gave one major band on gel electrophoresis (approx. 460 bp) which contained the desired sequence as demonstrated by blotting and hybridisation with the DIG-labelled central sequence previously isolated and nested PCR with primers 7 and 8, but was insufficiently pure for direct cloning and sequencing. These products were purified and used as the template for anchored PCR utilising the primer pairs 5 (5' end, degenerate) and 8 (specific to the 3' end of the known internal sequence) and 6 (3' end, degenerate) and 7 (specific to the 5' end of the known internal sequence) to amplify the gene in two fragments. PCR with these primers gave single products of sizes 276 bp (primers 5 and 8: 95°C, 30 s; 50°C, 60 s; 65°C, 60 s; for 30 cycles) and 305 bp (primers 6 and 7: 95°C, 30 s; 55°C, 60 s; 65°C, 60 s; for 30 cycles) as anticipated.

Several clones containing these products were isolated of which six of each were sequenced as detailed above. The sequences of the two clones overlapped completely from primer 7 to 8 and gave the full sequence for the gene encoding rusticyanin

```
GGTACATTGGATACTACATGGAAAGAGGCGACGCTT
```

Gly Thr Leu Asp Thr Thr Thr Lys Lys Ala Thr Leu 12

CCCAGTGAAAGGCTATGGAGAACAGGCACGG 72

Pro Gln Val Lys Ala Met Leu Lys Asp Thr Gly 24

AAAGTGAGCTGACACTGTTACCTACATTGAAGCAAG 108

Lys Val Ser Gly Asp Thr Val Thr Tyr Ser Gly Lys 36

ACTGTACATGTCGCGGCGGCGGTGTCGCCGGGA 144

Thr Val His Val Val Ala Ala Val Leu Pro Gly 48

TTCCAGTTCGAGTTTAGAGTTATGACAAAAAG 180

Phe Pro Phe Pro Ser Phe Glu Val His Asp Lys Lys 60

AACCCGACCTGGAGATTTCCGGGCGGGAAGGCAGTG 216

Asn Pro Thr Leu Glu Ile Pro Ala Gly Ala Thr Val 72

GACGTGAACTTCAAATAACAAACAAAGGGATCCGT 252

Asp Val Thr Phe Ile Asn Thr Asn Lys Gly Phe Gly 84

CATAGTTTGACACATCAAAGGGCACTTTAT 288

His Ser Phe Asp Ile Thr Lys Gly Pro Pro Tyr 96

GCCGTTATGCGGGATGCAACTTGGCGGGCCACCGT 324

Ala Val Met Pro Val Ile Asp Pro Val Leu Ala Gly 108

ACAGGTTCAGCTCCGTTCCCCAAGATGGCAAGTTCC 360

Thr Gly Phe Ser Pro Val Pro Lys Asp Gly Lys Phe 120

GGATACGAGTTTACTTACCTGGCATTCCCGCGGCGG 396

Gly Tyr Thr Asp Phe Thr Thr His Pro Thr Ala Gly 132

ACTCATCTACTAGTATATCAGTACCGGGCGAGCC 432

Thr Tyr Tyr Thr Val Cys Gin Ile Pro Gly His Ala 144

GCCACCGTGTGGGCCAGATGTGTTGTAAG 465

Ala Thr Gly Met Phe Gly Lys Ile Val Val Lys 154

Fig. 1. The nucleotide sequence encoding the *T. ferrooxidans* rusticyanin gene. The deduced amino acid sequence is shown below the nucleotide sequence. The differences between the *Tli*- (shown first) and the *Taq*-derived sequences are as follows: position 24 A = G, 62 T = C, 65 A = G, 441 T = C.
as shown in Fig. 1. Subsequently, primers were synthesised to the known sequence for each end of the gene (primers 9 and 10) and used to amplify the entire gene directly from genomic DNA using the Tli DNA polymerase under conditions of higher stringency: 95°C, 30 s; 55°C, 60 s; 65°C, 60 s; for 30 cycles. A single product was formed which was then cloned into Bluescript as before. Tli DNA polymerase has an integral ‘proofreading’ function, therefore minimising any errors due to mismatching of base pairs. Sequence analysis of four clones revealed five differences between the sequences obtained with Tli and Taq as indicated in Fig. 1. The final sequence is shown in Fig. 1, with the putative transcribed amino acid sequence which is identical to that published by Ronk et al. [2] from protein sequencing.

The isolation of the structural gene encoding rusticyanin is the first step in a molecular biology programme aimed at elucidating the structural features which are responsible for its unique properties: that is the high acid stability and the highest redox potential in the family of single blue copper proteins. This may also provide further impetus to studies designed to enhance the commercial application of T. ferrooxidans in bioleaching processes [17].

Acknowledgements

We would like to thank Professor K. Barker, Vice-Chancellor of De Montfort University, for providing special research funds for this project. We (J.F.H. and S.S.H.) would also like to acknowledge helpful discussions with Professor Lennart Lundberg and Dr. Goran Karlsson.

References