Identification of a Thiosulfate Utilization Gene Cluster from the Green Phototrophic Bacterium *Chlorobium limicola*[†,‡]

F. Verté,[§] V. Kostanjevecki,[§] L. De Smet,[§] T. E. Meyer,[||] M. A. Cusanovich,[||] and J. J. Van Beeumen*[,§]

Department of Biochemistry, Physiology, and Microbiology, Laboratory of Protein Biochemistry and Protein Engineering, University of Gent, B-9000 Gent, Belgium, and Department of Biochemistry and Molecular Biophysics, University of Arizona, Tucson, Arizona 85721

Received July 6, 2001; Revised Manuscript Received November 29, 2001

ABSTRACT: *Chlorobium* is an autotrophic, green phototrophic bacterium which uses reduced sulfur compounds to fix carbon dioxide in the light. The pathways for the oxidation of sulfide, sulfur, and thiosulfate have not been characterized with certainty for any species of bacteria. However, soluble cytochrome *c*-551 and flavocytochrome *c* (FCSD) have previously been implicated in the oxidation of thiosulfate and sulfide on the basis of enzyme assays in *Chlorobium*. We have now made a number of observations relating to the oxidation of reduced sulfur compounds. (1) Western analysis shows that soluble cytochrome *c*-551 in *Chlorobium limicola* is regulated by thiosulfate, consistent with a role in the utilization of thiosulfate. (2) A membrane-bound flavocytochrome *c*−sulfide dehydrogenase (which is normally a soluble protein in other species) is constitutive and not regulated by sulfide as expected for an obligately autotrophic species dependent upon sulfide. (3) We have cloned the cytochrome *c*-551 gene from *C. limicola* and have found seven other genes, which are also presumably involved in sulfur metabolism and located near that for cytochrome *c*-551 (SoxA). These include genes for a flavocytochrome *c* flavoprotein homologue (SoxF2), a nucleotidase homologue (SoxB), four small proteins (including SoxX, SoxY, and SoxZ), and a thiol−disulfide interchange protein homologue (SoxW). (4) We have established that the constitutively expressed FCSD genes (*soxEF1*) are located elsewhere in the genome. (5) Through a database search, we have found that the eight thiosulfate utilization genes are clustered in the same order in the *Chlorobium tepidum* genome (www.tigr.org). Similar thiosulfate utilization gene clusters occur in at least six other bacterial species but may additionally include genes for rhodanese and sulfite dehydrogenase.

The species of *Chlorobiaceae* (the green sulfur bacteria) carry out anoxygenic photosynthesis in which reduced sulfur compounds (sulfide, sulfur, thiosulfate) serve as electron donors for the reduction of carbon dioxide (*1*, *2*). All species are strictly anaerobic and obligately phototrophic. Elemental sulfur, which is formed extracellularly during oxidation of sulfide and thiosulfate, is further oxidized to sulfate (*3*). Utilization of thiosulfate as a photosynthetic electron donor by *Chlorobiaceae* is restricted to the "*thiosulfatophilum*" biovars. The two subspecies of *Chlorobium limicola* and *Chlorobium vibrioforme*, forma *thiosulfatophilum*, show a number of phenomena related to thiosulfate that are not shared by the other *Chlorobiaceae*: (1) the use of thiosulfate as photosynthetic electron donor, (2) the formation of thiosulfate during phototrophic sulfide oxidation, (3) the disproportionation of elemental sulfur in the light but in the absence of CO_2, (4) the possession of high levels of thiosulfate sulfur transferase or rhodanese, and (5) the possession of soluble cytochrome *c*-551 (*3*). It is the latter property which is of particular interest to us because soluble cytochrome *c*-551 had been shown to be enzymatically implicated in the oxidation of thiosulfate (*4*, *5*).

The soluble cytochromes found in the photosynthetic green sulfur bacterium, *C. limicola* strain Pond Mud (NCIB 8346) (*6*), all appear to be involved in the oxidation of sulfur-containing electron donors. Flavocytochrome *c*, which consists of a 47 kDa flavoprotein subunit and an 11 kDa cytochrome c_4 subunit containing a single heme (*7*−*11*), has been demonstrated to catalyze the oxidation of sulfide (*12*, *13*). Although a number of *c*-type cytochromes can function as electron acceptors for sulfide oxidation catalyzed by flavocytochrome *c* in vitro, the in vivo acceptor appears to be cytochrome *c*-555 (*4*) which forms an electrostatically stabilized complex with the enzyme (*14*).

Cytochrome *c*-555, in addition to its likely role as an electron acceptor during the oxidation of sulfide, may also function as an acceptor of electrons from thiosulfate. The 80 kDa thiosulfate-oxidizing enzyme appears to utilize soluble cytochrome *c*-551 as the initial electron acceptor (*4*, *5*). This enzyme has no visible absorption, hence does not possess flavin or heme. The overall reaction is greatly accelerated by the participation of cytochrome *c*-555, which is not directly reduced by the enzyme. Complex formation has also been demonstrated between cytochromes *c*-555 and

[†] This work was supported by the Fund for Scientific Research-Flanders (G.0282.01) and by a grant from the National Institutes of Health (GM21277).

[‡] The nucleotide sequences reported in this paper have been submitted to the GenBank/EMBL Data Bank with accession numbers AY074395 and AY074396.

* Corresponding author. Phone: 32/92645109. Fax: 32/92645338. E-mail: jozef.vanbeeumen@rug.ac.be.

[§] University of Gent.

[||] University of Arizona.

10.1021/bi011404m CCC: $22.00
Published on Web 02/07/2002

c-551 (*14*), suggesting that the actual electron acceptor may be the cytochrome *c*-551−*c*-555 complex.

Chlorobium-soluble cytochrome *c*-551 is more restricted in its distribution than is cytochrome *c*-555. It has been found only in the two thiosulfate-utilizing strains that have been studied: Pond Mud (NCIB 8346) and Larsen (NCIB 8327) (*6*, *15*, *16*). Soluble cytochrome *c*-551 is isolated as a dimer (*6*) and is composed of identical subunits of about 32 kDa (*15*, *16*). The redox potential is 135−150 mV, and it is less water-soluble than other cytochromes (it precipitates at 30−60% saturation ammonium sulfate). We previously determined the amino acid sequence of soluble cytochrome *c*-551 from *C. limicola* (*17*) as part of our ongoing effort to characterize the enzymes involved in sulfur metabolism. We have now extended our study of sulfur metabolism by isolating and characterizing the genes for the soluble cytochrome *c*-551 and FCSD.[1]

MATERIALS AND METHODS

Bacterial Strains and Plasmids. C. limicola strain Tassajara (DSM 249) and strain Pond Mud (NCIB 8346) cells were prepared in a medium modified after Biebl and Pfennig (*18*). The culture was grown under anaerobic photosynthetic conditions (40 W tungsten lamp; 30 °C). Most of the work reported herein was with strain Tassajara, and strain Pond Mud was used only to test for strain differences in the cytochrome *c*-551 by PCR and western analysis. *Escherichia coli* strain XL1-blue (*19*) was grown in LB medium (*20*) and used to detect α complementation for pUC18 (Amerscham Pharmacia Biotech, Uppsala, Sweden) derivatives on LB plates supplemented with 80 μM IPTG and 32 μg/mL X-gal. *E. coli* strain MC1061 (*21*, *22*) was used as a host for cloning techniques. Antibiotics were added at concentrations of 200 μg mL^{-1} for carbenicillin and 10 μg mL^{-1} for chloramphenicol.

Genomic Cloning. Chlorobium genomic DNA was isolated as described by Brown et al. (*23*). Plasmid preparations were carried out by standard procedures (*20*). On the basis of the *Chlorobium* cytochrome *c*-551 amino acid sequence (*17*), a set of degenerate primers was designed. PCR primers were synthesized by Amersham Pharmacia Biotech. The forward primer had the sequence 5′-TAYCARAARYTIGTIGAYG-CIGA-3′, and the reverse primer the sequence 5′-TTYC-TISWIGCIGGICCRTTRTA-3′. The probe was amplified by PCR using Taq-polymerase (Amersham Pharmacia Biotech). The amplified fragment of 780 bp was cloned in the pGEM-T vector (Promega, Madison, WI) and labeled via PCR with digoxigenin-dUTP (Roche Diagnostics, Mannheim, Germany). Via southern hybridization, a *Bam*HI fragment of 2.4 kb and two *Hin*dIII fragments of 4 and 0.9 kb were identified. Two subgenomic libraries were constructed, one for the *Bam*HI fragment of 2.4 kb and one for the *Hin*dIII fragment of 4 kb (*20*). The detection and identification of transformants were done with the nonradioactive DIG-DNA detection system (Roche Diagnostics). Double-stranded plasmid DNA was sequenced using Dye Terminator Cycle Sequencing (Perkin-Elmer, Foster City, CA). The sequencing was started from both ends with the universal primers M13F and M13R (New England Biolabs, Beverly, MA) and was continued with specific primers created on the basis of the new sequence. Once the sequence of cytochrome *c*-551 and of some of the flanking genes was obtained, a search against the *Chlorobium tepidum* genome (www.tigr.org) was performed. We found that cytochrome *c*-551 was part of a larger gene cluster that included an upstream copy of flavocytochrome *c*. Because the *Bam*HI fragment of 2.4 kb and the *Hin*dIII fragment of 4 kb did not encode all of the thiosulfate utilization cluster of *C. limicola*, we constructed a PCR probe based on the sequence of the flavoprotein subunit of the FCSD isozyme of *C. tepidum* with the forward primer 5′-ATGGGAAATACGATTTCT-CGC-3′ and the reverse primer 5′-ACCCCAGGTGTCCT-GTGAAAT-3′. Southern hybridization on genomic DNA of *C. limicola* revealed a *Hin*dIII fragment of 2.6 kb. After cloning and sequencing, we were able to compose the full gene cluster.

On the basis of the complete amino acid sequence of the cytochrome subunit and the partial sequence of the flavoprotein subunit of flavocytochrome *c* (FCSD) from *C. limicola* (*9*), a set of degenerate primers was designed. The forward primer had the sequence 5′-TGYGCIGGITGYCAYGGIAC-3′ and codes for the heme-binding sequence CAGCHGT of the cytochrome subunit. The reverse primer had the sequence 5′-GCIGCRTGCCAYTCRATCA-3′ and codes for the amino acid sequence MIEWHSS in the flavoprotein subunit. The probe was amplified by PCR using Taq-polymerase (Amersham Parmacia Biotech). The amplified fragment of 993 bp was cloned in the pGEM-T vector (Promega) and labeled via PCR with digoxigenin-dUTP (Roche Diagnostics). Southern hybridization revealed a *Bam*H1 fragment of 6 kb and two *Hind*III fragments of 2.4 and 5 kb used to construct subgenomic libraries. All other procedures were the same as those for cytochrome *c*-551.

Preparation of Protein Fractions. Protein samples were subjected to reducing SDS−PAGE (*24*) to detect the presence of the heme group by peroxidase activity (*25*). Total protein concentration was determined according to the method of Bradford (*26*) using the Bio-Rad 500-0006 kit with BSA as a standard. A total of 20 μg of boiled protein was loaded per lane for SDS−PAGE. The BenchMark prestained protein ladder (Invitrogen, Faraday) was used as molecular weight marker.

Western Blotting and Spectral Analysis. Cytochrome *c*-551 from crude extracts of *C. limicola* strain Tassajara was detected in western analysis with polyclonal serum obtained against purified cytochrome *c*-551 from strain Pond Mud (Eurogentec, Seraing, Belgium). Detection was done by chemiluminescence of the antigens blotted on Hybond-ECL membranes using horseradish peroxidase conjugated secondary antibodies (Roche Diagnostics). Flavocytochrome *c* sulfide dehydrogenase from crude extracts of *C. limicola* strain Tassajara was detected in western analysis with polyclonal serum obtained against the purified flavoprotein subunit of FCSD (Eurogentec). The blot signals were quantified using the program Quantity One (Bio-Rad, Hercules) and are expressed as the sum of the intensities of the pixels times the surface area (INT × mm^2). Reduced and oxidized spectra of recombinant cytochrome *c*-551 were

[1] Abbreviations: IPTG, isopropyl-β-D-thiogalactoside; X-gal, 5-bromo-4-chloro-3-indolyl-β-D-galactoside; *ccm*, cytochrome *c* maturation; MS medium, minimal salt medium; LB, Luria-Bertani; PAGE, polyacrylamide gel electrophoresis; BSA, bovine serum albumin; orf, open reading frame; FCSD, flavocytochrome *c*−sulfide dehydrogenase.

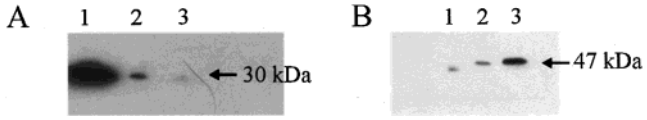


FIGURE 1: (A) Western blot analysis of protein fractions of *C. limicola* strain Tassajara with antibodies raised against cytochrome *c*-551 from strain PM: (1) periplasmic fraction, (2) cytoplasmic fraction, and (3) membrane fraction. (B) Western blot of protein fractions of *C. limicola* strain Tassajara with antibodies raised against flavocytochrome *c* from strain PM: (1) periplasmic fraction, (2) cytoplasmic fraction, and (3) membrane fraction.

measured on a Uvikon 943 spectrophotometer (Kontron, Milan, Italy).

RESULTS

Cellular Location of Cytochrome c-551 and Flavocytochrome c. It had previously been shown that the solubility of cytochrome *c*-551 is less than that of other cytochromes and that it precipitates at relatively low concentrations of ammonium sulfate (*6*), suggesting that it might actually be a membrane protein which is only partially solubilized during cell breakage. It was therefore essential to establish where the protein is located, whether it is truly soluble, cytoplasmic, periplasmic, or membrane bound. Periplasmic, cytoplasmic, and membrane fractions of *C. limicola* strain Tassajara were prepared, and protein blots were incubated with antisera raised against cytochrome *c*-551 (from strain Pond Mud). There was strong hybridization to the periplasmic fraction, weak reaction to the cytoplasmic fraction, and almost no response with the membrane fraction, as shown in Figure 1A. Thus, cytochrome *c*-551 is a soluble periplasmic protein in both strains. However, we cannot exclude the possibility that it interacts in vivo with another protein, such as SoxX.

Antibodies were raised against the flavoprotein subunit of FCSD (from strain PM) and western blots performed with extracts from strain Tassajara. In contrast to cytochrome *c*-551, FCSD is located primarily in the membrane, with significantly less protein in the soluble fractions, as shown in Figure 1B. In fact, a slight reaction with the soluble fraction may be due to residual membranes that had not completely sedimented upon centrifugation. Although it was unexpected to find FCSD in the membrane fraction of strain Tassajara because of the soluble nature of the same protein from strain PM, the FCSD from *Ectothiorhodospira vacuolata* is also located in the membrane (*27*).

Regulation of Cytochrome c-551 and FCSD. All strains and species of *Chlorobium* use sulfide and, in fact, require sulfide to maintain anaerobiosis and also to use it as a reductant for carbon dioxide fixation. Thiosulfate, on the other hand, is only used by some representatives of the green bacteria to fix CO_2, and it cannot replace sulfide as a reductant. The genes for thiosulfate utilization even reside on a plasmid in some strains (*28*). One would therefore expect that thiosulfate utilization might be regulated. Although it had been shown that cytochrome *c*-551 has a role in thiosulfate oxidation (*4*, *5*), it had also been reported that cytochrome biosynthesis was not regulated (*16*). However, we have now shown through western blots that cytochrome *c*-551 can be regulated under certain conditions. *C. limicola* strain Tassajara was grown in a basal medium, as described by Biebl and Pfennig (*18*), with various combinations of acetate, sulfide, and thiosulfate. For those cultures in which thiosulfate was included in the medium, cytochrome *c*-551 was produced approximately four times more than when thiosulfate was omitted, as shown in Figure 2A. The lack of regulation previously reported for stationary phase cultures (*16*), may have been due to the intermediate formation of thiosulfate during growth on sulfide, as found for all species that utilize thiosulfate in addition to sulfide. Thus, both enzyme assays and regulation of cytochrome *c*-551 by thiosulfate point to its involvement in the oxidation of thiosulfate.

When the same experiment was performed with antibodies against FCSD, it was clear that it is not regulated by either sulfide or thiosulfate, as shown in Figure 2B. The lack of regulation of FCSD by sulfide in strain Tassajara is somewhat surprising because it has been implicated in sulfide oxidation and is regulated in *E. vacuolata* (*27*). On the other hand, sulfide is required for growth of *Chlorobium* but not *E. vacuolata*; thus, the enzymes for oxidation of sulfide should be constitutive in *Chlorobium*.

Cloning the Cytochrome c-551 Gene. The reported amino acid sequence of the soluble cytochrome *c*-551 from *C. limicola* contains 258 residues, a single heme, and a disulfide bond in the 30 kDa protein monomer (*17*). At the outset, we assumed that all strains of *C. limicola* would be similar if not identical. However, this proved not to be the case, as shown here. Nevertheless, the published amino acid sequence was used to design degenerate primers, and a PCR fragment of the appropriate size and sequence was obtained from *C. limicola*, strain Tassajara. This was used, in turn, for southern analysis, resulting in the isolation of a 2.4 kb *Bam*HI fragment and a *Hin*dIII fragment of 4 kb, which were completely sequenced as shown in Figure 3. An upstream 2.6 kb *Hin*dIII fragment was subsequently obtained and sequenced as well.

Cytochrome *c*-551 from strain Tassajara is encoded by *soxA* and contains a total of 286 amino acid residues, including a 21-residue signal peptide directing it to the periplasmic space, in agreement with the location established by western blots. The mature protein thus contains 265 residues, which sums up to a molecular mass of 30 365 Da. The initiator ATG codon is preceded by a normal GGAG ribosome binding site. However, the sequence is not identical to that previously determined for the purified protein from the same nominal strain, differing at 41 positions (84% identity). We have found only one copy of the cytochrome *c*-551 gene (*soxA*) in *Chlorobium*, suggesting that the protein previously sequenced had been derived from another strain or species than had been reported (*17*). On the basis of a subsequent analysis of strain Pond Mud (NCIMB 8346), which showed that the protein encoded by the PCR product was identical to the published protein sequence (*17*), the previous result is likely to correspond to strain Pond Mud rather than to strain Tassajara, which was being grown in the laboratory at the same time. Thus, the SoxA from these two strains is as different as can be expected for any two species of *Chlorobium*.

Identification of a Thiosulfate Utilization Gene Cluster. Cytochrome *c*-551 is currently believed to be the electron acceptor for the thiosulfate oxidizing enzyme rather than the enzyme itself (*4*, *5*). Thus, it is to be expected that the cytochrome *c*-551 gene might be associated with additional

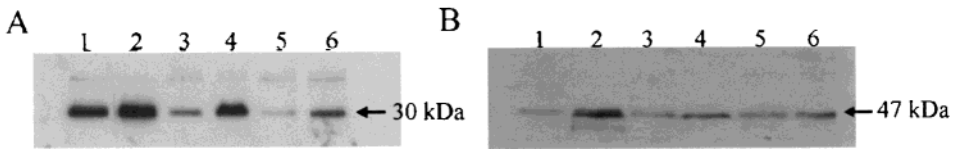


FIGURE 2: (A) Western blot analysis of *C. limicola* strain Tassajara periplasmic protein fraction with antibodies raised against cytochrome *c*-551 from strain PM. The minimal medium was supplemented with (1) sulfide + thiosulfate, (2) sulfide + thiosulfate + acetate, (3) sulfide + acetate, (4) thiosulfate + acetate, (5) sulfide, and (6) thiosulfate. The cells were harvested after 3 days (exponential phase). The relative staining intensities (INT × mm^2) are 7.5, 9.6, 2.1, 5.0, 1.0, and 3.8, respectively. (B) Western blot of *C. limicola* strain Tassajara membrane protein fraction with antibodies raised against flavocytochrome *c* from strain PM. The minimal medium was supplemented with (1) sulfide + thiosulfate, (2) sulfide + thiosulfate + acetate, (3) sulfide + acetate, (4) thiosulfate + acetate, (5) sulfide, and (6) thiosulfate. The cells were harvested after 3 days (exponential phase). The relative staining intensities (INT × mm^2) are 1.0, 4.9, 1.8, 3.3, 2.7, and 3.0, respectively.

genes in the pathway. That this is so can be shown by analysis of the genes flanking the cytochrome *c*-551 gene (*soxA*). We were able to locate eight open reading frames on the combined *Bam*H1 and *Hin*dIII fragments in the following order: *soxF2*, *soxX*, *soxY*, *soxZ*, *soxA*, *orf106*, *soxB*, and *soxW*, as shown in Figures 3 and 4. Gene names were chosen by analogy to those of *Paracoccus*, which we subsequently found to be similar. There are differences in the number of genes in the two clusters and in the gene order, but the comparison gives an idea of the minimal set of genes required for thiosulfate utilization. In *Chlorobium*, all but *soxZ* are preceded by signal peptides and thus appear to be periplasmic. SoxZ is probably translocated in tandem with SoxY by analogy with the hydrogenase subunits, only one of which has a signal peptide (*29*). Four genes appear to lack a ribosome binding site, including *soxF2*, *orf106*, *soxB*, and *soxW*. It is possible that the ATGs we have chosen are not the initiator codons, but all four sequences have the double Arg or Lys motif near the N terminus, which is frequently observed with periplasmic signal peptides. There is a string of T's between *soxF2* and *soxX*, *soxX* and *soxY*, and *soxB* and *soxW*, which suggests that the genes preceding and following these strings may not be translated equally. There are hypothetical promoter regions in front of *soxF2*, *soxY*, and *soxZ*.

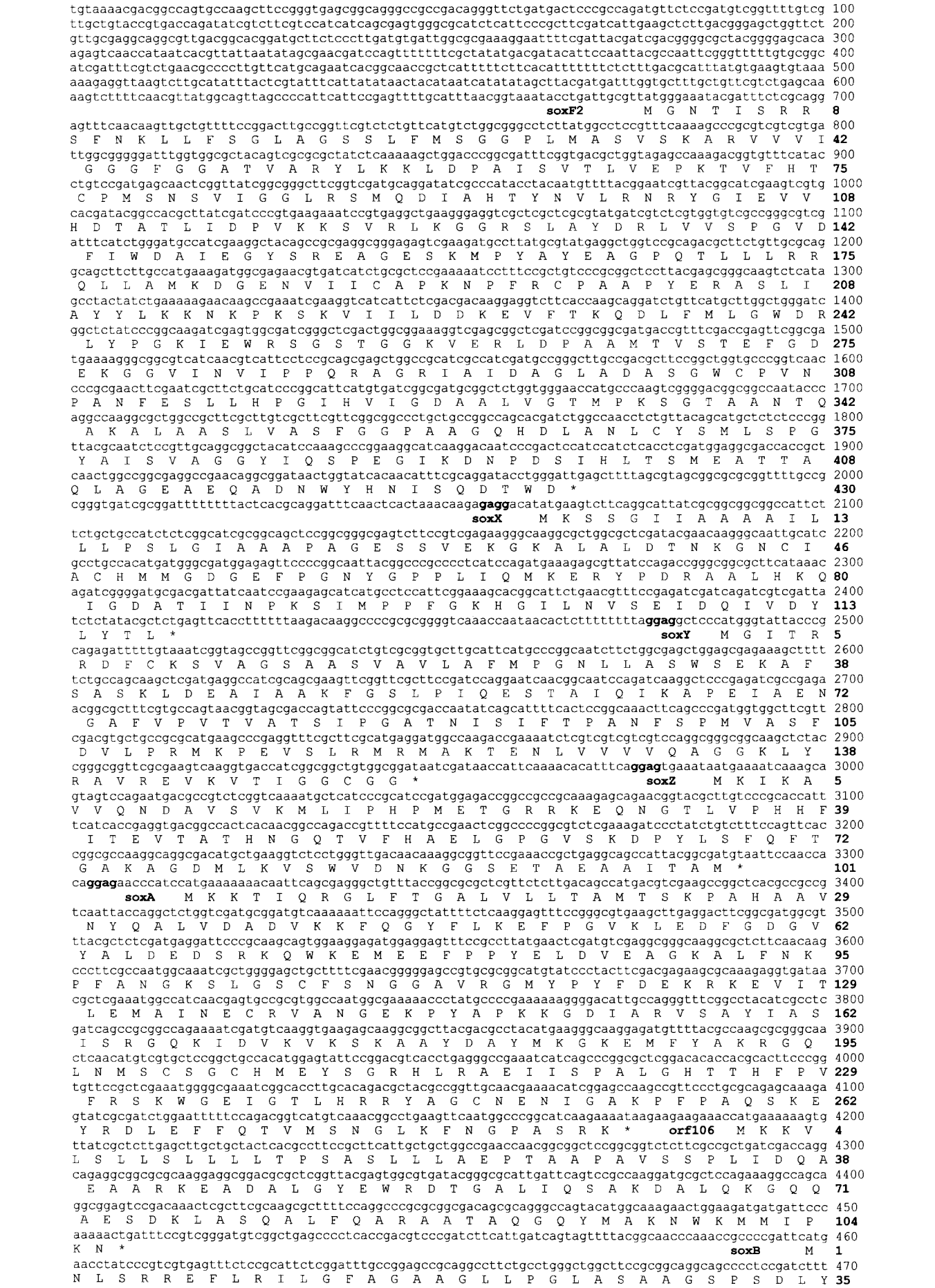
Cloning of the Flavocytochrome c (FCSD) Genes. The previously reported amino acid sequence of the FCSD cytochrome subunit and the partial sequence of the flavoprotein subunit (*9*) was used to design degenerate oligonucleotide primers which were used for PCR with strain Tassajara. A fragment of the appropriate size and sequence was obtained and used for southern analysis of chromosomal DNA. Two *Hin*dIII fragments of 2.4 and 5 kb were cloned and sequenced, as shown in Figure 5. The cytochrome subunit gene (*soxE*) is followed by the gene for the flavoprotein subunit (*soxF1*), as expected. However, there are no other genes for electron-transfer proteins that might function in sulfur metabolism. Upstream of *soxEF1* in the same orientation, there is a gene for a lipoprotein A of 136 residues. Downstream of *soxEF1*, and in the opposite orientation, there is a gene coding for a 154-residue orf that contains a CXXC motif generally characteristic of thioredoxins. Further downstream, there is a 462-codon orf that shows homology to membrane-spanning multidrug resistance transporter proteins. The translated sequence of the monoheme cytochrome subunit, SoxE, is not the same as the one that has been reported previously from strain Tassajara (*9*), but they are 71% identical. A PCR product from strain PM does correspond to the published sequence, consistent with the likelihood that the strains used in the protein sequence studies had been mislabeled. The flavoproteins, SoxF1, from the two strains are 76% identical.

Both the cytochrome and flavoprotein genes contain leader sequences apparently directing the proteins to the periplasmic space. The gene for the flavoprotein contains a normal leader, which suggests that it is cleaved following translocation across the membrane. However, the cytochrome gene contains a leader that does not have a normal cleavage site. The leader sequence is followed by a 15-residue segment, rich in Ala and Pro, suggestive of a flexible arm that is sometimes present in peripherally membrane-bound proteins and which allows them to swing around in order to react with different reaction partner proteins. This result is consistent with the membrane location of FCSD indicated by western blots. In contrast to *Chlorobium*, the protein from *E. vacuolata* is membrane bound via an uncleaved signal peptide in the flavoprotein subunit (*27*).

DISCUSSION

Thiosulfate Utilization Gene Cluster. Prior to this study, it had been shown that cytochrome *c*-551 is involved in thiosulfate oxidation, but it was uncertain as to whether it was actually soluble, whether it was located in the periplasmic space, as is the case with the majority of *c*-type cytochromes, and whether it was regulated. We have now shown that it is a soluble, periplasmic protein, that the gene contains a leader sequence directing the protein to the periplasm, and that it is regulated by thiosulfate. Furthermore the cytochrome *c*-551 gene is associated with a cluster of seven other genes; we believe that they may be involved in sulfur metabolism in general and in thiosulfate utilization in particular. Consistent with our observations with *C. limicola*, the *C. tepidum* genome sequence (www.tigr.org) contains a single cytochrome *c*-551 gene which is actually closer to strains Pond Mud and Tassajara sequences than they are to one another, as shown in Figure 6 (28 and 32 differences (11−12%)). It is not surprising that *C. tepidum* should have a cytochrome *c*-551 because it is also capable of utilizing thiosulfate (*30*). The eight genes presumed to be involved in thiosulfate utilization in *C. limicola* strain Tassajara are conserved in the same order in *C. tepidum*, thus strengthening this viewpoint.

With the presence of a similar cluster of eight genes in two species of *Chlorobium* established, we examined other bacterial species in data banks for *sox* genes. There exist at least seven relatively close homologues of the *Chlorobium* soluble cytochrome *c*-551, namely, from *Paracoccus pan-*


soxF2
soxX
soxY
soxZ
soxA
orf106
soxB

```
atgatctcggtcagtccggcgacattcgcttgttgcacatcacggatacccatgcccagctcatgccgatctactaccgtgaacccagcctgaacctcgg 4800
  D  L  G  Q  S  G  D  I  R  L  L  H  I  T  D  T  H  A  Q  L  M  P  I  Y  Y  R  E  P  S  L  N  L  G  68
gctcggccaggcattcggacgtcctcctcacctcgtgacggagtcgctcttgaaatattacggcatcgcacctggtacgccgctcgcacatgcctatacc 4900
 L  G  Q  A  F  G  R  P  P  H  L  V  T  E  S  L  L  K  Y  Y  G  I  A  P  G  T  P  L  A  H  A  Y  T  101
gcgatcaactatgccgaggcggcgcagcggttcggaaaggttggcggtttcgcgcatctgaagacgctcgttgaccggatgcgttcggagtacggatccg 5000
A  I  N  Y  A  E  A  A  Q  R  F  G  K  V  G  G  F  A  H  L  K  T  L  V  D  R  M  R  S  E  Y  G  S  D  135
acaagaccctcctgctcgatggcggcgacacctggcagggttcgggaacagcgttctggaatcgcggcatggacatggtcgaggcctgcaatttgctcgg 5100
  K  T  L  L  L  D  G  G  D  T  W  Q  G  S  G  T  A  F  W  N  R  G  M  D  M  V  E  A  C  N  L  L  G  168
cgtcgatgtgatgacgggtcactgggagttcacctatctcgaagaggaggtactcaaaaacctcgcggctttcaagggcgacttcgtggcgcagaatatc 5200
 V  D  V  M  T  G  H  W  E  F  T  Y  L  E  E  E  V  L  K  N  L  A  A  F  K  G  D  F  V  A  Q  N  I  201
aaggtcaaggaggatgcgctcttcaacggagccaaggcgttcgacgagaattccggacacgccttccgtccctacgtggtcaaagcggtgggcaagcacc 5300
K  V  K  E  D  A  L  F  N  G  A  K  A  F  D  E  N  S  G  H  A  F  R  P  Y  V  V  K  A  V  G  K  H  R  235
gcgtcgccgtgattggccaggcgtttccctacacgccgatcgccaatcccgcccgcttcattccgaactggaccttcggcatcaatgccagcgacatgca 5400
  V  A  V  I  G  Q  A  F  P  Y  T  P  I  A  N  P  A  R  F  I  P  N  W  T  F  G  I  N  A  S  D  M  Q  268
gcagcttgtcgataccgttcgctccaagaagaagcctgacgcggtcgtgctgatctcgcacaacgggatggatgttgacgtcaagctcgcgcaggtggtc 5500
 Q  L  V  D  T  V  R  S  K  K  K  P  D  A  V  V  L  I  S  H  N  G  M  D  V  D  V  K  L  A  Q  V  V  301
agcggcatcgacgtgatttttcggcggccacacccatgacggcgtaccgcagcccttcgtcgtgcagaacgccaaggggcgaacgctcgtcaccaacgccg 5600
S  G  I  D  V  I  F  G  G  H  T  H  D  G  V  P  Q  P  F  V  V  Q  N  A  K  G  R  T  L  V  T  N  A  G  335
gatcgaacggcaaattcctcggcgtcatcgacctcaagctcggcaatggcggcgtcaaggagtataattacaagttgcttcccgtcttttccaacgaact 5700
  S  N  G  K  F  L  G  V  I  D  L  K  L  G  N  G  G  V  K  E  Y  N  Y  K  L  L  P  V  F  S  N  E  L  368
tccggcgcacaacgggatgcaggcgctgatcgacaaaacccgcgcaccctacctcgacaagctcaacgagccactcgccgtggccggatcgctgctctac 5800
 P  A  H  N  G  M  Q  A  L  I  D  K  T  R  A  P  Y  L  D  K  L  N  E  P  L  A  V  A  G  S  L  L  Y  401
cggcgcggcaatttcgacggcccgttcgaccagatcatctgcaacgcgcttcgccagcagaacgacgcgcagatttcgctttcgcccggcttccgctggg 5900
R  R  G  N  F  D  G  P  F  D  Q  I  I  C  N  A  L  R  Q  Q  N  D  A  Q  I  S  L  S  P  G  F  R  W  G  435
gtacgagcattttgcccggccagaccatcacgatggagcacgtgctcgaccagacctgcatgacctatcccgaaacctacgtgcgcgacatgaccggcca 6000
  T  S  I  L  P  G  Q  T  I  T  M  E  H  V  L  D  Q  T  C  M  T  Y  P  E  T  Y  V  R  D  M  T  G  Q  468
gcagatcaaggacattctcgaagacgtggccgacaacccttttcaatctcgatcccttctaccagcagggcggcgacatggtgcgcaccggcggtctgagc 6100
 Q  I  K  D  I  L  E  D  V  A  D  N  L  F  N  L  D  P  F  Y  Q  Q  G  G  D  M  V  R  T  G  G  L  S  501
tatcggatcgatccgatggcctcgatgggcaagcgcatcgacaacatgcggctcgaaaacggcaaggtggtggatgcttcgcaaaagtaccgtgttgcag 6200
Y  R  I  D  P  M  A  S  M  G  K  R  I  D  N  M  R  L  E  N  G  K  V  V  D  A  S  Q  K  Y  R  V  A  G  535
gctgggcgaccgtgggcgcaaagtcgccgggcgagccggtttgggataccgtggccgcttatctgaaagataaaaaggtggtcgaagtgaaaaagctcaa 6300
  W  A  T  V  G  A  K  S  P  G  E  P  V  W  D  T  V  A  A  Y  L  K  D  K  K  V  V  E  V  K  K  L  N  568
ccagcccgaattcaaaaacatgggcagcaatccgggaatcgatctcacctgaggcgaagatcgtcagcggacaaggctgttgcagaagcgaaaccgcgac 6400
 Q  P  E  F  K  N  M  G  S  N  P  G  I  D  L  T  *                                                  584
tgattcccgtttctgtcctgaacggagggctggaatgtgcacccttcgtcagccaacagatttttgcttttgagaactccgtttgcttttttttctgaaag 6500
agcctgttgacggacttcgtttgttttgaaaagctcaatccatgaatggttcatgaaatccattaaaaaattgttcgccacctcaatcgtggcgctttcg 6600
                                    soxW    M  K  S  I  K  K  L  F  A  T  S  I  V  A  L  S  16
atgattggcgcgcttccggcaacaagcgcgagcgcggcggttcctcccgcgagccaggccgtcgcgggcaaagtcgccccggcgttccggatcaaaacgc 6700
M  I  G  A  L  P  A  T  S  A  S  A  A  V  P  P  A  S  Q  A  V  A  G  K  V  A  P  A  F  R  I  K  T  L  50
tcgacggcaaggagttgaagagttcgcagcttgccggacggccctacatcgtcaacttttcgcctcgtggtgcccgccctgccgtgaggagctgcccgg 6800
  D  G  K  E  L  K  S  S  Q  L  A  G  R  P  Y  I  V  N  F  F  A  S  W  C  P  P  C  R  E  E  L  P  G  83
catggtggcattgcagaaaaagtatgcaaacaagggcttcacttttatcggcatcgcttttcgggatcgccccgcgacgctgcccgattttctctgggag 6900
 M  V  A  L  Q  K  K  Y  A  N  K  G  F  T  F  I  G  I  A  F  R  D  R  P  A  T  L  P  D  F  L  W  E  116
atgggggtcgactatccggtcggcctcaccactcctgaactcgaagccgcgctttggcaagctcatgccgggcggaaaaatccgcgcgattccggcaacct 7000
M  G  V  D  Y  P  V  G  L  T  T  P  E  L  E  A  A  F  G  K  L  M  P  G  G  K  I  R  A  I  P  A  T  F  150
tcgtcgttggccgcgatggcaagattctcaatgccgtgagtggcggcctcaccagagaggacttcgagtcgctcatcatcaaggcggtcaacaccagacc 7100
  V  V  G  R  D  G  K  I  L  N  A  V  S  G  G  L  T  R  E  D  F  E  S  L  I  I  K  A  V  N  T  R  P  183
tgtaaagtaagctcacgaagcgaggcttcccggatttccctctccctcgaccctgcccgaacatcctctccatctcctccattcgctgaaccggcccaga 7200
 V  K  *                                                                                             185
gagagcgccggttcagcgattcttgctttatgcgtggcgatcatttgcagccggccatcacggagctgtcaggcaaaatggcttttccgcgccggcgctg 7300
tttcttgtttggtttttgtagttcatattaaatatattcatttgtttgataaacgatttggttatatagttaaatcgttgattttaaggcagtgacaggtt 7400
tctgtttccaaggctgtttgggcgttatgcgcgcaggctgtgcgcggagagctgaaaaagctt                                       7463
```

FIGURE 3: DNA sequence and translated amino acid sequence of the thiosulfate utilization gene cluster of *C. limicola* strain Tassajara. The ribosome binding sites are boldfaced.

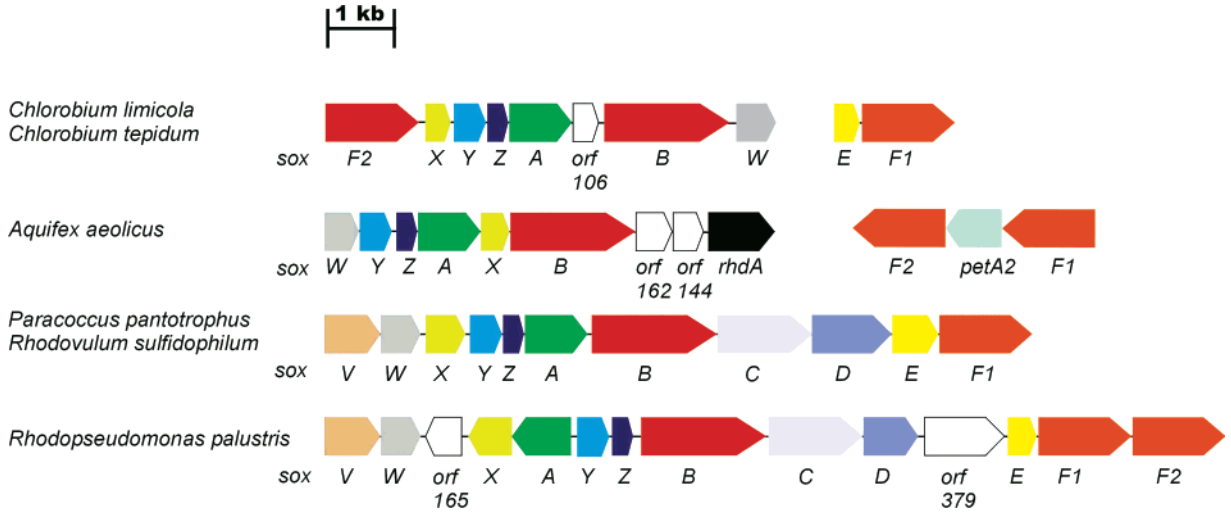
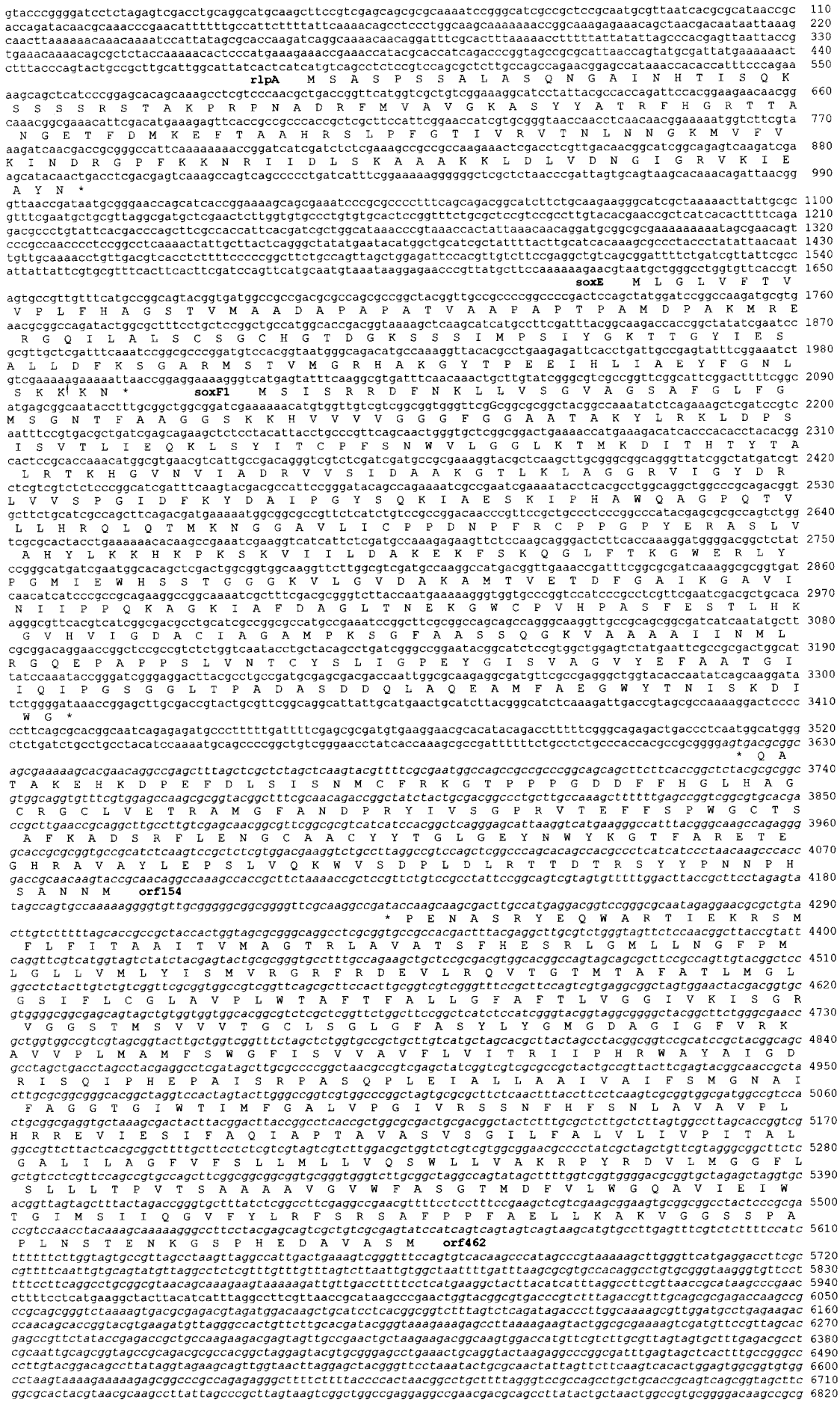


FIGURE 4: DNA sequence and translated amino acid sequence of the flavocytochrome *c* operon from *C. limicola* strain Tassajara.

totrophus (*31*−*33*), *Aquifex aeolicus* (*34*), *Rhodopseudomonas palustris*, *Alcaligenes eutrophus* (now called *Ralstonia metallidurans*), *Magnetococcus* sp. (www.jgi.doe.gov), *Thiobacillus* sp. (*35*), and *Rhodovulum sulfidophilum* (*36*), as shown in Figure 6. Some of these genome sequences are incomplete, yet we have located several other genes that belong to the *sox* cluster, which lead us to believe that all of the genes will eventually be found. The *Rhodopseudomonas* and *Ralstonia* SoxA proteins have one heme-binding site like that of *Chlorobium*. The *Paracoccus*, *Magnetococcus*, *Thio-*


rlpA
soxE
soxF1
orf154
orf462

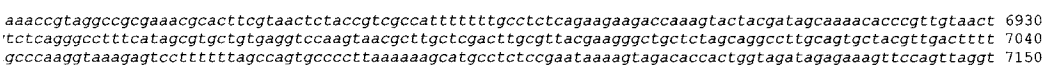


FIGURE 5: Composition of the thiosulfate utilization gene clusters of *Chlorobium* (this work and www.tigr.org), *A. aeolicus* (*34*), *P. pantotrophus* (*31*, *32*, *50*), *Rps. palustris* (www.jgi.doe.gov), and *R. sulfidophilum* (*36*).

bacillus, *Rhodovulum*, and *Aquifex* SoxA sequences differ in having two apparent heme binding sites. *Aquifex* gene AQ1807 was unidentified in the published genome sequence (*34*) but is in fact a homologue of cytochrome *c*-551 or SoxA, with about 24% identity. *Aquifex* is not known to utilize thiosulfate. However, there are similarities in the genetic context of the cytochromes as components of related gene clusters in these species, which strengthens their structural and functional assignment to SoxA, as will be shown below. Furthermore, SoxA is regulated by thiosulfate in *Rhodovulum* (*36*).

There are no conserved methionines or histidines that could fill the role of sixth ligand to the heme in SoxA, although it has recently been reported that cysteine is the ligand to both hemes in *R. sulfidophilum* (*37*). *Chlorobium* SoxA has only one heme and three conserved cysteines, two of which (78 and 109) form a disulfide (*17*). Therefore, the free Cys219 may be the sixth ligand to the single heme. The two cysteines that form a disulfide in *Chlorobium* are likely to be bound to the other heme in *Rhodovulum* and all diheme SoxA. It is also apparent from examination of the amino acid sequences in Figure 6 that motifs characteristic of N-terminal and C-terminal helices of class I c-type cytochromes are present in SoxA, which would indicate that the two hemes originated from gene doubling. For example, there is a conserved Gly X X X aromatic motif 9−10 residues in front of the first heme binding site and 10 residues ahead of the second heme, which indicates a 6−7 residue insertion between the N-terminal helix and the heme in both domains. Following the sixth ligand cysteine by 21−30 residues in the first domain and 21 residues in the second, there is a conserved aromatic-hydrophobic pair of residues, characteristic of the C-terminal helix.

There are weak homologues of SoxA in *Bordetella bronchiseptica* (www.sanger.ac.uk), *Shewanella oneidensis*, gene 3254 (www.tigr.org), *Synechocystis* sp. 6803, gene SLL1359 (*38*), *Wolinella succinogenes* (Arnold, B., and Simon, J., unpublished work; Genbank accession number AJ318778), *Campylobacter jejuni* (*39*), and *Pseudomonas aeruginosa*, genes 2481 and 4571 (*40*) (sequences not shown) that appear to have a sixth ligand cysteine in the first domain and a more normal methionine in the second domain. These genes do not appear to be located in the same *sox* cluster as for *Chlorobium*, *Paracoccus*, *Aquifex*, *Rhodovulum*, and *Rhodopseudomonas*. Therefore, it is not possible to tell whether they may be functionally related to *soxA*.

Downstream of the *C. limicola* cytochrome *c*-551 gene (*soxA*), there is a large open reading frame encoding a 585-residue protein, SoxB, which is 88% identical to that found in *C. tepidum* (www.tigr.org). There are a number of relatively close homologues of SoxB, from *Paracoccus* (43% identical) (*32*, *41*), *Rhodopseudomonas* (41% identical) (www.jgi.doe.gov), and *Aquifex* (35% identical) (*34*), as shown in Figure 7. *Ralstonia* (www.jgi.doe.gov), *Magnetococcus* (www.jgi.doe.gov), *Rhodovulum* (*36*), *Thiobacillus novellus* (*42*), and *Methylobacterium extorquens* (M. E. Lidstrom, University of Washington) plus a number of other species (*43*) (not shown) are equally closely related. However, there are a large number of more distant homologues which generally function as nucleotidases. The three-dimensional structure of the *E. coli* 5′-nucleotidase is known to have a binuclear cluster of zinc atoms at the active site, which serves to bind the terminal phosphate of the substrate as well as the hydrolytic water (*44*). A catalytic His-Asp pair is assumed to effect hydrolysis of the phosphate. When SoxB is aligned with the *E. coli* nucleotidase (not shown), five out of six metal ligands, plus the catalytic His-Asp pair, are conserved (see Figure 7). We can thus postulate that SoxB is responsible for hydrolysis of the terminally bound *S*-sulfonate resulting from the oxidation of thiosulfate and sulfur (see also ref *33*).

Chlorobium soxW, *soxX*, *soxY*, *soxZ*, and *orf106* are present in the *C. tepidum* genome (www.tigr.org) in the same order as in strain Tassajara. They are 82%, 88%, 90%, 84%, and 80% identical, respectively. There are also homologues of *soxW*, *soxX*, *soxY*, and *soxZ* upstream of *soxA* in *Paracoccus* (*31*−*34*, *45*) and *Rhodovulum* (*36*). SoxY and SoxZ are 33% and 30% identical to those of *Chlorobium*, respectively, as shown in Figure 8. Homologues of SoxYZ are present in the preliminary genome sequence of *Rhodopseudomonas*, but *soxX* and *soxA* appear to be divergently transcribed between *soxW* and *soxYZ* (www.jgi.doe.gov). There are also homologues of *soxY*, and *soxZ* in *Aquifex* upstream of *soxA*, labeled AQ1810 and AQ1809 (*34*), which are 36% and 38% identical to those of *Chlorobium*, respectively, but the *soxX* or AQ1806 is located between *soxA* and *soxB*. We found *soxY* in the unfinished genomes of *Magnetococcus* and *Alcaligenes* but did not locate *soxZ*. A homologue of orf106 has not yet been found in *Aquifex*, *Rhodopseudomonas*, *Rhodovulum*, or *Paracoccus*, and in fact, we have not found a homologue in any species. Therefore, it may not have a role in sulfur metabolism. *SoxW* is homologous to the thiol−disulfide interchange proteins that are distant relatives of thioredoxin. The nearest homologue of *soxW* appears to be *helX* from *Rhodobacter capsulatus*, which is thought to be involved in the reduction of the cysteines in apocytochromes *c* prior to heme attachment (*46*). *Paracoccus* has a *soxW* homologue upstream of *soxY* also called *orf3* or *shxW* (*31*). *Aquifex* also has a related thiol−disulfide interchange protein (AQ1811) just upstream of the *soxY* homologue (AQ1810). SoxW contains a CXXC motif characteristic of thioredoxin and protein disulfide reductases. The thiol−disulfide redox balance in the periplasm has been shown to influence *c*-type cytochrome biogenesis in *E. coli* (*47*, *48*). The finding of an operon that encodes both the structural gene of a cytochrome *c* and a homologue of the disulfide interchange proteins suggests that the latter is involved in the maturation process rather than being more directly involved in sulfur metabolism.

In *C. tepidum*, a gene was found to occur upstream of *soxX*, encoding a flavoprotein (SoxF2) homologous to flavocytochrome *c* (www.tigr.org). We subsequently cloned

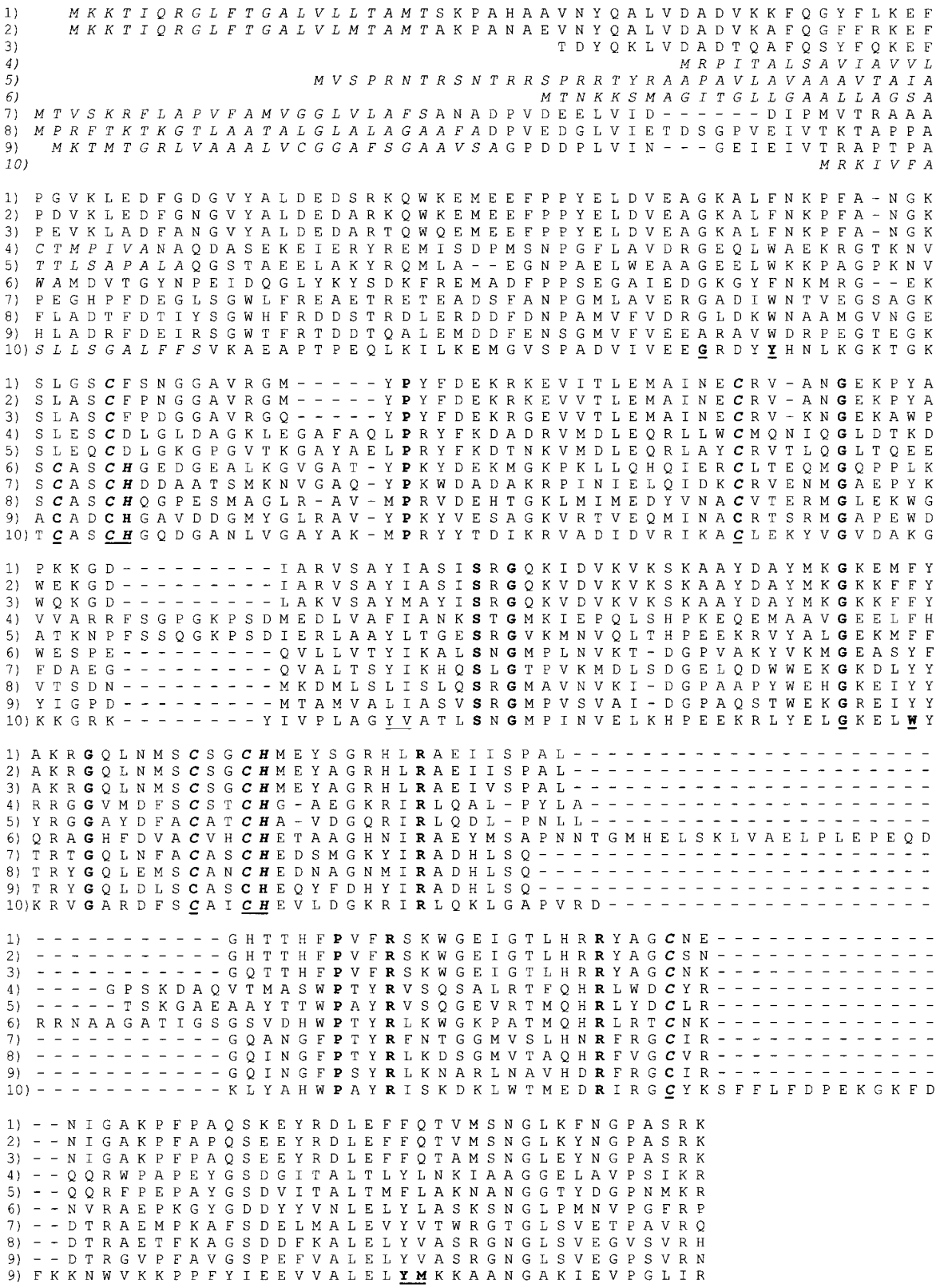


FIGURE 6: Alignment of cytochrome *c*-551 (SoxA) from (1) *C. limicola* strain Tassajara (this work), (2) *C. tepidum* (www.tigr.org), (3) *C. limicola* presumably strain Pond Mud (*17*; this work), (4) *Rps. palustris* (www.jgi.doe.gov), (5) *A. eutrophus* (www.jgi.doe.gov), (6) *Magnetococcus* sp. (www.jgi.doe.gov), (7) *Thiobacillus* sp. (*35*), (8) *P. pantotrophus* strain GB17 (*31*, *32*), (9) *R.* sulfidophilum (*36*), and (10) *A. aeolicus* (gene AQ1807; *34*). Conserved residues are boldfaced; conserved cysteines, heme binding sites, and helical motifs are underlined; and the leader sequences are italicized.

a 2.6 kb *Hin*dIII fragment encoding this gene in *C. limicola*. SoxF2 is not the same flavocytochrome as has been characterized as a soluble protein in *Chlorobium* species, including *C. limicola* and *C. tepidum*, but is more divergent (58% identical as opposed to 86% identity between the orthologous pairs). Unlike FCSDs, there is no gene for a cytochrome subunit upstream of the SoxF2 flavoprotein gene. The cytochrome gene that follows the flavoprotein in the

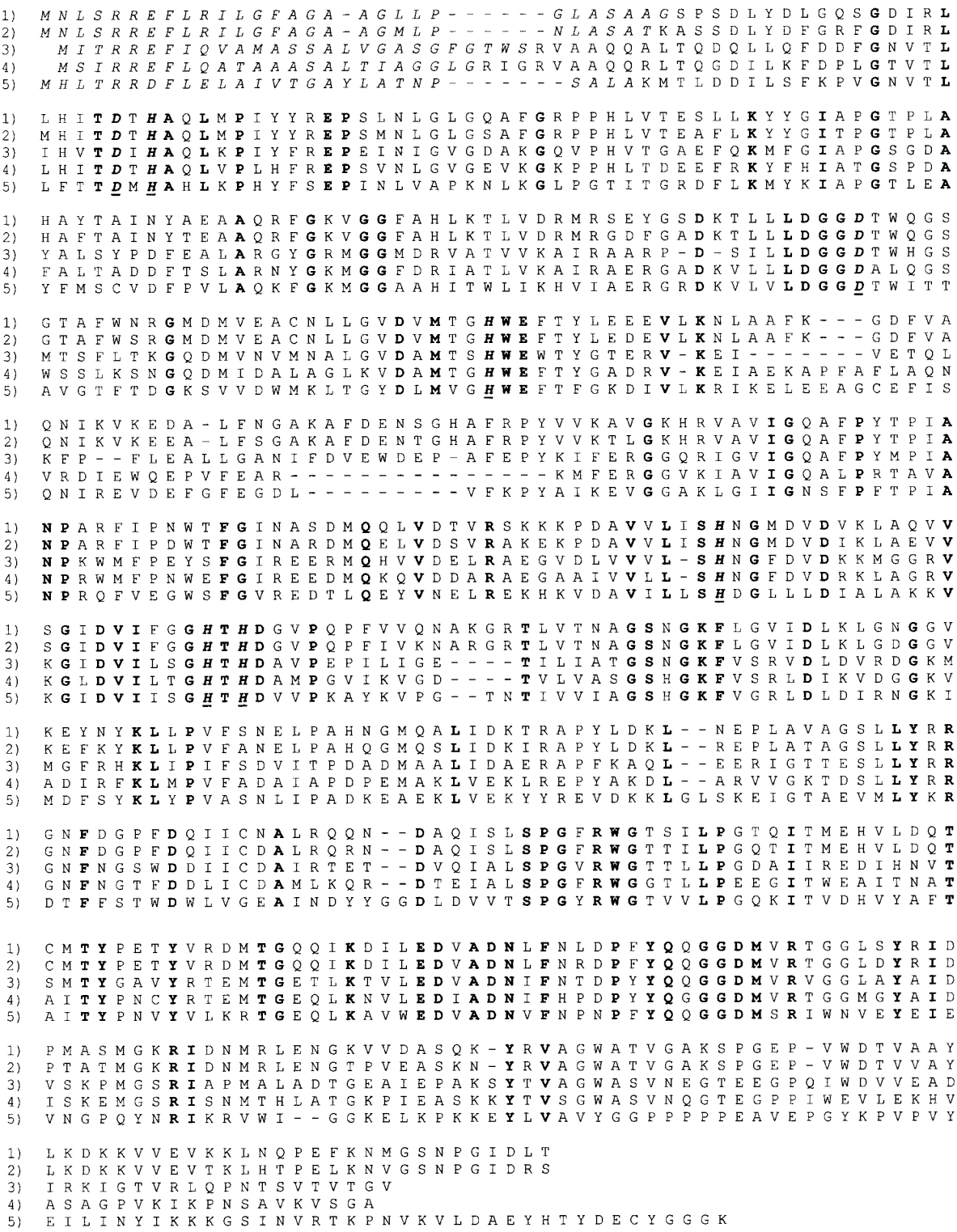


FIGURE 7: Alignment of soxB proteins from (1) *C. limicola* strain Tassajara (this work), (2) *C. tepidum* (www.tigr.org), (3) *P. pantotrophus* (*31*), (4) *Rps. palustris* (www.jgi.doe.gov), and (5) *A. aeolicus* (*34*). Conserved residues are boldfaced, the metal binding and catalytic residues are underlined, and the leader sequences are italicized.

thiosulfate gene cluster appears to be slightly more closely related to *soxX* than to the FCSD cytochrome subunit *soxE*, although it is not possible to tell at this time whether the supposed *SoxX* interacts with *SoxA* or with *SoxF2* because of the weak sequence homology to either protein (SoxX homologues are aligned in ref *36*). The genes for the well-characterized FCSD in *C. tepidum* are located elsewhere on the chromosome and appear to be transcribed independently. Likewise, we have cloned the more commonly observed FCSD from strain Tassajara, which is distantly located on the chromosome relative to the thiosulfate cluster. A rare lipoprotein A (RLPA) gene occurs upstream of FCSD in both species, but *C. tepidum* has a gene for NADH oxidase between RLPA and FCSD. To distinguish the two flavoproteins, we have designated the gene for the flavoprotein in the *sox* cluster as *soxF2* and those for the well-known FCSD as *soxEF1*. *SoxF2* is relatively closely related to *SoxF1,* as shown in the alignment of Figure 9. Most importantly, it

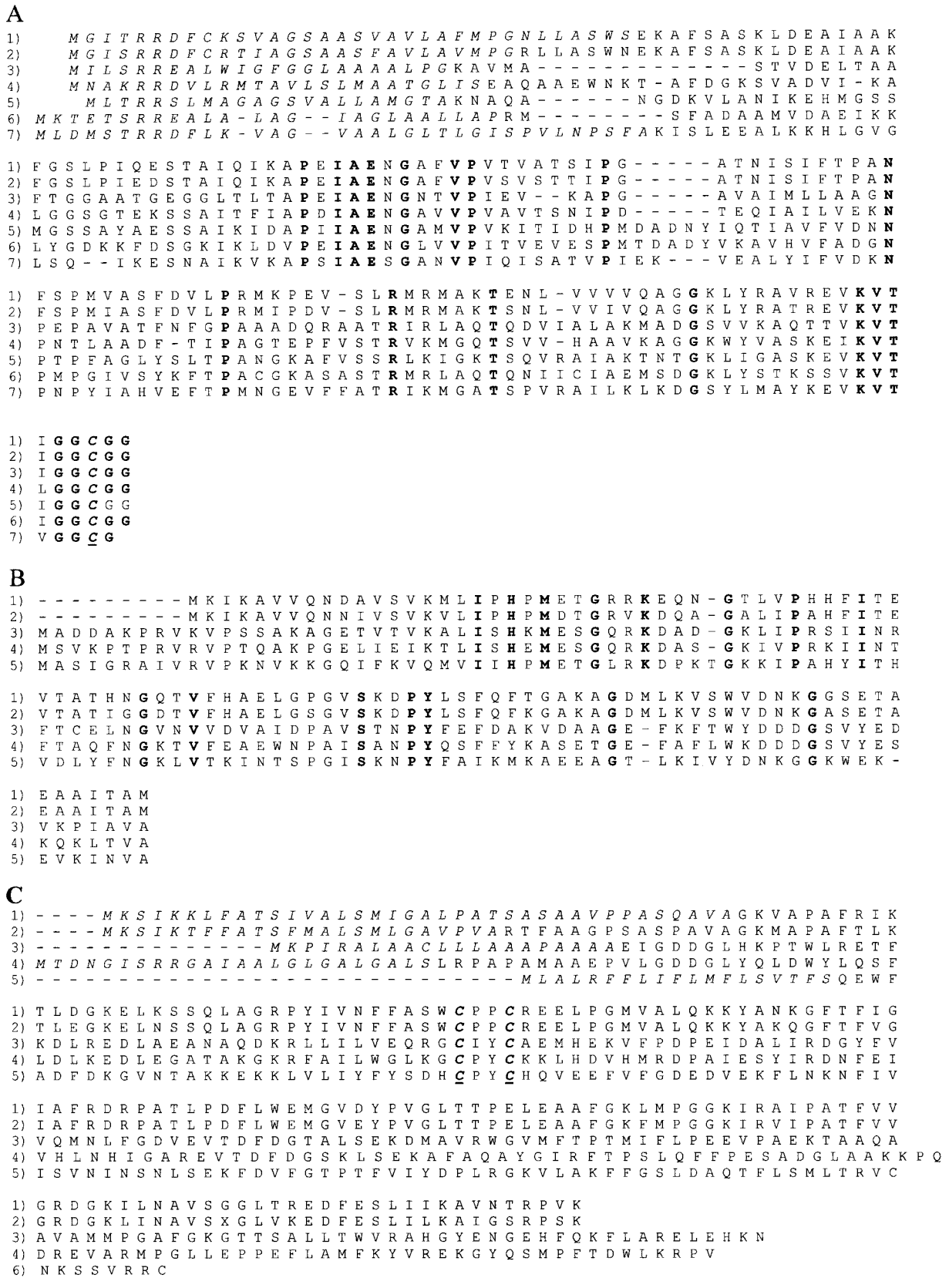


FIGURE 8: (A) Alignment of SoxY proteins from (1) *C. limicola* strain Tassajara (this work), (2) *C. tepidum* (www.tigr.org), (3) *P. pantotrophus* GB17 (*31*), (4) *A. eutrophus* (www.jgi.doe.gov), (5) *Magnetococcus* sp. (www.jgi.doe.gov), (6) *Rps. palustris* (www.jgi.doe.gov), and (7) *A. aeolicus* gene product AQ1810 (*34*). Conserved residues are boldfaced, the conserved cysteine is underlined, and leader sequences are italicized. (B) Alignment of SoxZ proteins from (1) *C. limicola* strain Tassajara (this work), (2) *C. tepidum* (www.tigr.org), (3) *P. pantotrophus* GB17 (*31*), (4) *Rps. palustris* (www.jgi.doe.gov), and (5) *A. aeolicus* gene product AQ1809 (*34*). Conserved residues are boldfaced. (C) Alignment of SoxW proteins from (1) *C. limicola* strain Tassajara (this work), (2) *C. tepidum* (www.tigr.org), (3) *P. pantotrophus* GB17 (*31*), (4) *Rps. palustris* (www.jgi.doe.gov), and (5) *A. aeolicus* gene product AQ1811 (*34*). The active site cysteines are boldfaced and underlined and the leader sequences are italicized.

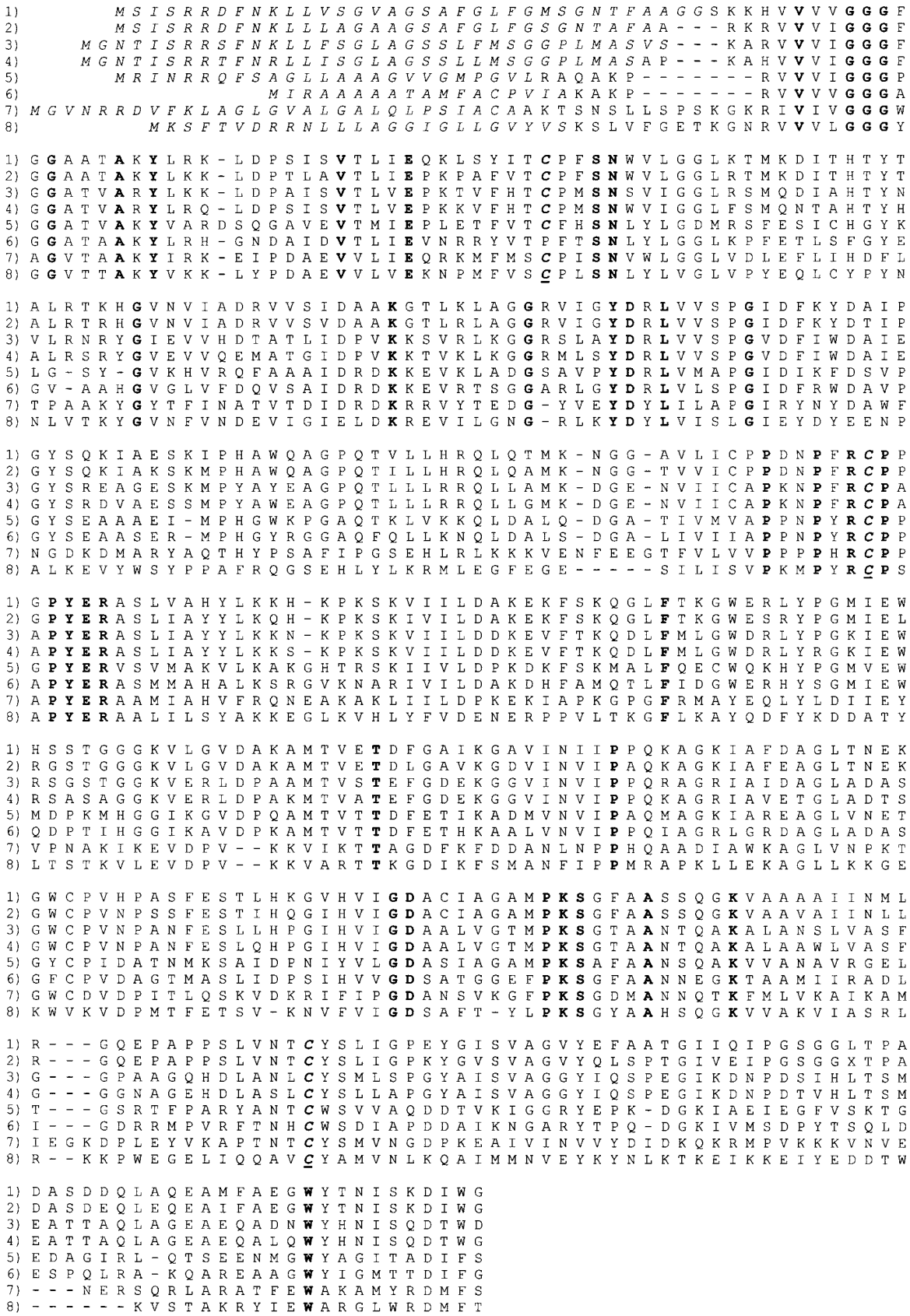


FIGURE 9: Alignment of flavocytochrome *c* flavoprotein subunits and homologues: (1) *C. limicola* strain Tassajara SoxF1 (this work), (2) *C. tepidum* SoxF1 (www.tigr.org), (3) *C. limicola* strain Tassajara SoxF2 (this work), (4) *C. tepidum* SoxF2 (www.tigr.org), (5) *Rps. palustris* SoxF1 (www.jgi.doe.gov), (6) *Rps. palustris* SoxF2 (www.jgi.doe.gov), (7) *A. aeolicus* SoxF1 (*34*), and (8) *A. aeolicus* SoxF2 (*34*).

has the Cys161−Cys337 disulfide (*A. vinosum* numbering) that is characteristic of FCSD and sulfide quinone reductase. However, the apparent lack of a cytochrome subunit raises the question as to what is the electron acceptor.

FCSD is thought to function as a sulfide dehydrogenase (*8*), but when the genes were inactivated in *Allochromatium vinosum*, there was no phenotype (*49*). If, on the other hand, there is generally more than one of the FCSD-like flavoprotein genes, such as SoxF1 and SoxF2 in the *C. tepidum* and *C. limicola* genomes, we would expect there to be no phenotype unless both sets of genes were inactivated. The fact that there was no phenotype can also be due to the presence of alternate sulfide-oxidizing enzymes. *A. vinosum* extracts catalyze electron transfer from sulfide to externally added duroquinone, indicating the presence of the enzyme sulfide:quinone oxidoreductase (SQR) (*49*). There is a somewhat more distant homologue of FCSD in the *C. tepidum* genome that has the requisite disulfide and might correspond to an SQR, and the absence of a leader sequence is consistent with this tentative identification. In *Paracoccus*, *soxEF* genes occur downstream of the *soxAB* genes, but the corresponding proteins do not have sulfide dehydrogenase activity, and knockout mutants have no phenotype (*50*). There are two genes for periplasmic FCSD-like flavoproteins in *Aquifex*, AQ232 and AQ235, but they appear to be associated with a copy of the Rieske iron−sulfur protein gene rather than with cytochrome subunit genes, and they are located elsewhere in the chromosome with respect to the *sox* genes (*34*). It remains to be seen whether a small cytochrome gene, AQ1806, in the *Aquifex* conserved *sox* gene cluster encodes a subunit of these FCSD flavoproteins or is actually a SoxX. *Aquifex* also has a homologue of SQR (AQ2186) that is apparently involved in the oxidation of sulfide (*51*). Thus, there are three related flavoproteins potentially involved in sulfide oxidation in both *Chlorobium* and *Aquifex*. *Rhodovulum* has an FCSD homologue, and *Rhodopseudomonas* also contains a pair of adjacent FCSD-like flavoproteins as part of the *sox* gene cluster. Unlike *Chlorobium* and *Aquifex*, the *Rhodovulum* and *Rhodopseudomonas* genes are preceded by a cytochrome subunit.

Comparison with the Thiosulfate Utilization Gene Cluster of Paracoccus. The genes and enzymes involved in thiosulfate utilization have been more thoroughly characterized from *Paracoccus* than from *Chlorobium*, as reviewed by Friedrich and co-workers (*33*, *45*) and Kelly et al. (*52*). Nevertheless, it is still unknown how thiosulfate is oxidized in *Paracoccus* or how many genes may be involved in this or in any other species. In view of the genetic similarities for thiosulfate utilization in the aforementioned species and shown in Figure 4, we will discuss the *Paracoccus* system here. *Paracoccus versutus* (formerly *Thiobacillus versutus*) (*53*) uses a multienzyme system to oxidize thiosulfate to sulfate. Previous enzymatic studies have shown that it minimally contains two colorless proteins and two *c*-type cytochromes, located in the periplasmic space of the organism (*54*, *55*). They are called enzyme A (16 kDa), enzyme B (63 kDa), cytochrome *c*-552.5 (56 kDa; subunit 29 kDa), and cytochrome *c*-551 (260 kDa; subunit 43 kDa). Enzyme A is now known to be a mixture of 16 and 12 kDa proteins, and the former has been established as a thiosulfate binding protein (*33*, *45*, *56*).

A 13kb genomic region of *P. pantotrophus* GB17 (formerly *Thiosphaera pantotropha*) is involved in lithotrophic thiosulfate oxidation. The coding region for sulfur (actually thiosulfate) oxidation (sox) was identified by the isolation of a transposon Tn5-mob mutant with a sox-minus phenotype (*57*). A 6.3 kb subgenomic region contains six open reading frames, *soxABCDEF*, as described previously (*32*, *41*). An additional upstream 4.3 kb fragment encodes orf1, orf2, and *soxVWXYZA* genes (*30*). The remainder of *soxF* plus *soxGH* were found further downstream (*50*). The N-terminal protein sequence of *Paracoccus* cytochrome *c*-552.5 indicates that it is encoded by *soxX* and *soxA* (*31*). Enzyme A is composed of *SoxY* and *SoxZ*. The C terminus of SoxY is very highly conserved and contains a cysteine which is likely to be the site of binding of sulfur, sulfite, and thiosulfate (*56*). Enzyme B of the thiosulfate-oxidizing system of *Paracoccus* is encoded by *soxB* (*32*, *41*). The mature translated *soxC* exhibits high amino acid sequence similarity to one domain of a eukaryotic molybdoenzyme, sulfite oxidase, which also contains a cytochrome b_5 domain. The *Paracoccus* enzyme was designated sulfite dehydrogenase, emphasizing the absence of the b_5 domain, although it is now thought to be a sulfur oxidizing enzyme (*33*). A mutant, GbsoxC (carrying an in-frame deletion in *soxC*), was impaired in its ability to oxidize thiosulfate and was unable to grow lithoautothrophically with thiosulfate but grew well with nitrate as a nitrogen source or as an electron acceptor (*32*). *SoxD* codes for a 44 kDa periplasmic diheme cytochrome c_4c_2 hybrid protein of 384 amino acids that is a subunit of the purified sulfite dehydrogenase and appears to serve as electron acceptor for SoxC in lieu of the missing cytochrome b_5 domain (*58*). SoxCD may be unique to *Paracoccus*, *Rhodovulum*, and *Rhodopseudomonas* because we were unable to locate *soxC* or *soxD* genes in either the *Chlorobium* or the *Aquifex* genomes. *SoxE* codes for a 26 kDa periplasmic diheme cytochrome c_2c_4 hybrid protein of 236 amino acids (in which the order of the two domains is reversed). *SoxF* codes for a 420 residue protein (*50*) which is 37% identical to the sequence of the flavoprotein subunit of FCSD from *Allochromatium vinosum*. However, SoxF did not have sulfide dehydrogenase activity, it was not necessary for oxidation of sulfide, and deletion of the SoxF gene did not impair growth with thiosulfate (*50*).

CONCLUSION

We have identified a periplasmic thiosulfate utilization gene cluster in *Chlorobium* species that is analogous to the more thoroughly characterized system in *Paracoccus*. We have also demonstrated by homology that similar systems exist in *Rhodopseudomonas*, *Rhodovulum*, and *Aquifex* and are likely to be present in *Magnetococcus* and *Ralstonia*. The thiosulfate utilization system appears to consist of five proteins which are universally present and essential for activity, including a thiosulfate, sulfur, and sulfite binding protein (SoxYZ), a cytochrome *c* complex (SoxXA) which may be the site of oxidation of sulfur compounds bound to SoxYZ, and a "nucleotidase" (SoxB) which may hydrolyze the oxidized SoxYZ−*S*-sulfonate complex. The roles of the flavocytochrome *c*−sulfide dehydrogenase (SoxEF, or SoxF) homologues, rhodanese, and a thioredoxin homologue (SoxW) are less clear. In *Chlorobium* and *Rhodovulum*, SoxA is regulated by thiosulfate. In *Paracoccus* and presumably in *Rhodovulum* and *Rhodopseudomonas*, sulfite dehydrogenase (SoxCD), in addition to the aforementioned proteins, is essential for the oxidation of thiosulfate as well as sulfite but is not present in *Chlorobium* or *Aquifex*.

ACKNOWLEDGMENT

F.V. is a Research Assistant of the IWT (Vlaams Instituut voor de Bevordering van het Wetenschappelijk-Technologisch Onderzoek in de Industrie, Flanders). We thank Drs. D. Bryant, U. Fischer, C. Friedrich, C. Dahl, and J. Cole for helpful discussions.

REFERENCES

1. Pfennig, N. (1989) in *Bergey's Manual of Systematic Bacteriology* (Staley, J. T., Bryant, M. P., Pfennig, N., and Holt, J. C., Eds.) pp 1682−1709, Williams and Wilkins, Baltimore, MD.
2. Pfennig, N., and Trüper, H. G. (1992) in *The Prokaryotes* (Starr, M. P., Stolz, H., Trüper, H. G., and Schlegel, H. G., Eds.) pp 3583−3592, Springer-Verlag, Berlin, Germany.
3. Trüper, H. G., Lorenz, C., Schedel, M., and Steinmetz, M. (1988) in *Green Photosynthetic Bacteria* (Olson, J. M., Ormerod, J. G., Amesz, J., Stackebrandt, E., and Trüper, H. G., Eds.) pp 189−200, Plenum Press, New York.
4. Kusai, A., and Yamanaka, T. (1973) *Biochim. Biophys. Acta 325*, 304−314.
5. Kusai, A., and Yamanaka, T. (1973) *Biochem. Biophys. Res. Comm. 51*, 107−112.
6. Meyer, T. E., Bartsch, R. G., Cusanovich, M. A., and Mathewson, J. H. (1968) *Biochim. Biophys. Acta 153*, 854−861.
7. Bartsch, R. G., Meyer, T. E., and Robinson, A. R. (1968) in *Structure and Function of Cytochromes* (Okunuki, K., Kamen, M. D., and Sekuzu, I., Eds.) pp 443−451, University Park Press, Baltimore, MD.
8. Cusanovich, M. A., Meyer, T. E., and Bartsch, R. G. (1991) in *Chemistry and Biochemistry of Flavoenzymes* (Muller, F., Ed.) pp 377−399, CRC Press, Boca Raton, FL.
9. Van Beeumen, J., Van Bun, S., Meyer, T. E., Bartsch, R. G., and Cusanovich, M. A. (1990) *J. Biol. Chem. 265*, 9793−9799.
10. Yamanaka, T. (1976) *J. Biochem. 79*, 655−660.
11. Yamanaka, T., Fukumori, Y., and Okunuki, K. (1979) *Anal. Biochem. 95*, 209−213.
12. Kusai, A., and Yamanaka, T. (1973) *FEBS Lett. 34*, 235−237.
13. Yamanaka, T., and Fukumori, Y. (1980) in *Flavins and Flavoproteins* (Yagi, K., and Yamano, T., Eds.) pp 631−639, Japan Scientific Societies Press, Tokyo.
14. Davidson, M. W., Meyer, T. E., Cusanovich, M. A., and Knaff, D. B. (1986) *Biochim. Biophys. Acta 850*, 396−401.
15. Fischer, U. (1988) in *Green Photosynthetic Bacteria* (Olson, J. M., Ormerod, J. G., Amesz, J., Stackebrandt, E., and Trueper, H. G. Eds.) pp 127−131, Plenum Press, New York.
16. Steinmetz, M., and Fischer, U. (1982) *Arch. Microbiol. 131*, 19−26.
17. Klarskov, K., Verté, F., Van Driessche, G., Meyer, T. E., Cusanovich, M. A., and Van Beeumen, J. (1998) *Biochemistry 37*, 10555−10562.
18. Biebl, H., and Pfennig, N. (1978) *Arch. Microbiol. 117*, 9−16.
19. Bullock, W. O., Fernandez, J. M., and Short, J. M. (1987) *BioTechniques 5*, 376−378.
20. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) Molecular cloning: A Laboratory Manual, 2nd ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
21. Meissner, P. S., Sisk, W. P., and Berman, M. L. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 4171−4175.
22. Wertman, K. F., Wyman, A. R., and Botstein, D. (1986) *Gene 49*, 253−262.
23. Brown, J. W., Haas E. S., James, B. D., Hunt, D. A., and Pace, N. R. (1991) *J. Bacteriol. 173*, 3855−3863.
24. Laemmli, U. K. (1970) *Nature 227*, 680−685.
25. Thomas, P. E., Ryan, D., and Ledwin, W. (1976) *Anal. Biochem. 75*, 168−176.
26. Bradford, M. M. (1976) *Anal. Biochem. 72*, 248−254.
27. Kostanjevecki, V., Brige, A., Meyer, T. E., Cusanovich, M. A., Guisez, Y., and Van Beeumen, J. J. (2000) *J. Bacteriol. 182*, 3097−3103.
28. Mendez-Alvarez, S., Pavon, V., Esteve, I., Guerrero, R., and Gaju, N. (1995) *FEMS Microbiol. Lett. 134*, 279−285.
29. Rodrigue, A., Chanal, A., Beck, K., Mueller, M., and Wu, L. F. (1999) *J. Biol. Chem. 274*, 13223−13228.
30. Wahlund, T. M., Woese, C. R., Castenholz, R. W., and Madigan, M. T. (1991) *Arch. Microbiol. 156*, 81−90.
31. Friedrich, C. G., Quentmeier, A., Bardischewsky, F., Rother, D., Kraft, R., Kostka, S., and Prinz, H. (2000) *J. Bacteriol. 182*, 4677−4687.
32. Wodara, C., Bardischewsky, F., and Friedrich, C. G. (1997) *J. Bacteriol. 179*, 5014−5023.
33. Friedrich, C. G., Rother, D., Bardischewsky, F., Quentmeier, A., and Fischer, J. (2001) *Appl. Environ. Microbiol. 67*, 2873−2882.
34. Deckert, G., Warren, P. V., Gaasterland, T., Young, W. G., Lenox, A. L., Graham, D. E., Overbeek, R., Snead, M. A., Aujay, M., Huber, R., Feldman, R. A., Short, J. M., Olsen, G. J., and Swanson, R. V. (1998) *Nature 392*, 353−358.
35. Mukhopadhyaya, P. N., Deb, C., Lahiri, C., and Roy, P. (2000) *J. Bacteriol. 182*, 4278−4287.
36. Appia-Ayme, C., Little, P. J., Matsumoto, Y., Leech, A. P., and Berks, B. C. (2001) *J. Bacteriol. 183*, 6107−6118.
37. Cheesman, M. R., Little, P. J., and Berks, B. C. (2001) *Biochemistry 40*, 10562−10569.
38. Kaneko, T., and Tabata, S. (1997) *Plant Cell Physiol. 38*, 1171−1176.
39. Parkhill, J., Wren, B. W., Mungall, K., Ketley, J. M., Churcher, C., and Basham, D. (2000) *Nature 403*, 665−668.
40. Stover C. K., et al. (2000) *Nature 406*, 959−964.
41. Wodara, C., Kostka, S., Egert, M., Kelly, D. P., and Friedrich, C. G. (1994) *J. Bacteriol. 176*, 6188−6191.
42. Kappler, U., Friedrich, C. G., Truper, H. G., and Dahl, C. (2001) *Arch. Microbiol. 175*, 102−111.
43. Petri, R., Podgorsek, L., and Imhoff, J. F. (2001) *FEMS Microbiol. Lett. 197*, 171−178
44. Knofel, T., and Strater, N. (1999) *Nat. Struct. Biol. 6*, 448−453.
45. Friedrich, C. G. (1998) *Adv. Microb. Physiol. 39*, 235−289.
46. Beckman, D. L., and Kranz, R. G. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 2179−2183.
47. Crooke, H., and Cole, J. (1995) *Mol. Microbiol. 15*, 1139−1150.
48. Sambongi, Y., and Ferguson, S. J. (1994) *FEBS Lett. 353*, 235−240.
49. Reinartz, M., Tschäpe, J., Brüser, T., Trüper, H. G., and Dahl, C. (1998) *Arch. Microbiol. 170*, 59−68.
50. Rother, D., Henrich, H. J., Quentmeier, A., Bardischewsky, F., and Friedrich, C. G. (2001) *J. Bacteriol. 183*, 4499−4508.
51. Nubel, T., Klughammer, C., Huber, R., Hauska, G., and Schutz, M. (2000) *Arch. Microbiol. 173*, 233−244.
52. Kelly, D. P., Shergill, J. K., Lu, W. P., and Wood, A. P. (1997) *Antonie Van Leeuwenhoek 71*, 95−107.
53. Katayama, Y., Hiraishi, A., and Kuraishi, H. (1995) *Microbiology 141*, 1469−1477.
54. Lu, W.-P. (1986) *FEMS Microbiol. Lett. 34*, 313−317.
55. Lu, W.-P., and Kelly, D. P. (1984) *J. Gen. Microbiol. 130*, 1683−1692.
56. Quentmeier, A., and Friedrich, C. G. (2001) *FEBS Lett. 503*, 168−172.
57. Chandra, T. S., and Friedrich, C. G. (1986) *J. Bacteriol. 166*, 446−452.
58. Quentmeier, A., Kraft, R., Kostka, S., Klockenkamper, R., and Friedrich, C. G. (2000) *Arch. Microbiol. 173*, 117−125.

BI011404M