# Cobalamin (vitamin $B_{12}$) biosynthesis: identification and characterization of a *Bacillus megaterium cobI* operon

Evelyne RAUX*[1], Anne LANOIS†, Martin J. WARREN*, Alain RAMBACH‡ and Claude THERMES†

*Department of Molecular Genetics, Institute of Ophthalmology, UCL, 11-43 Bath Street, London EC1V 9EL, U.K., †Centre de Génétique Moléculaire, Laboratoire associé à l'Université Pierre et Marie Curie, Centre National de la Recherche Scientifique, 91198 Gif sur Yvette, France, and ‡CHROMagar, 267 rue Lecourbe, 75015 Paris, France

A 16 kb DNA fragment has been isolated from a *Bacillus megaterium* genomic library and fully sequenced. The fragment contains 15 open reading frames, 14 of which are thought to constitute a *B. megaterium* cobalamin biosynthetic (*cob*) operon. Within the operon, 11 genes display similarity to previously identified *Salmonella typhimurium* cobalamin biosynthetic genes (*cbiH*$_{60}$, *-J*, *-C*, *-D*, *-ET*, *-L*, *-F*, *-G*, *-A*, *cysG*$^{A}$ and *btuR*), whereas three do not (*cbiW*, *-X* and *-Y*). The genes of the *B. megaterium* *cob* operon were compared with the cobalamin biosynthetic genes of *Pseudomonas denitrificans*, *Methanococcus jannaschii* and *Synechocystis* sp. Taking into account the presence of *cbiD* and *cbiG*, the absence of a *cobF*, *cobG* and *cobN*, *-S* and *-T*, it was concluded that *B. megaterium*, *M. jannaschii* and *Synechocystis* sp., like *S. typhimurium*, synthesize cobalamin by an anaerobic pathway, in which cobalt is added at an early stage and molecular oxygen is not required.

## INTRODUCTION

Cobalamins (vitamin $B_{12}$) represent a class of one of Nature's most complex small molecules that are unique among the vitamins in that they are made exclusively by micro-organisms [1]. There are two main coenzyme forms of vitamin $B_{12}$, methylcobalamin and adenosylcobalamin [2,3], associated with methylation and rearrangement reactions. These are required by enzymes such as glutamate mutase, ribonucleotide reductase, methionine synthase and methylmalonyl-CoA mutase.

Cobalamins belong to the same prosthetic group family as haem, chlorophyll, sirohaem and coenzyme $F_{430}$ and are all derived from the common macrocyclic biosynthetic intermediate uroporphyrinogen III [4]. However, from an evolutionary standpoint, vitamin $B_{12}$ can be considered the matriarch of the family through the evolutionary ancient role of adenosylcobalamin in ribonucleotide reductase [5–7]. There is therefore an interest not only in how cobalamin is biosynthesized but also in how the pathway evolved, especially with relation to the emergence of the pathways for the other modified tetrapyrroles including haem and chlorophyll.

The nomenclature used to describe the genes of cobalamin biosynthesis in *Pseudomonas denitrificans* and *Salmonella typhimurium* is quite confusing. For instance, in *P. denitrificans* all the genes required for cobalamin biosynthesis are prefixed *cob*, whereas in *S. typhimurium* the genes required for cobinamide biosynthesis are prefixed *cbi* and the genes required to convert cobinamide into cobalamin are termed *cob*. Thus the *cobA* genes in *P. denitrificans* and *S. typhimurium* refer to quite separate enzymes. For clarity, where duplication of gene names occurs, we have added a superscript to identify the organism. Furthermore, in *P. denitrificans*, *cobI* refers to the C-20 methyltransferase, whereas in *S. typhimurium cobI* refers to all the genes of the *cob* operon responsible for cobinamide biosynthesis. The *S. typhimurium* equivalent of *Escherichia coli* BtuR is CobA, whereas the *P. denitrificans* equivalent is called CobO.

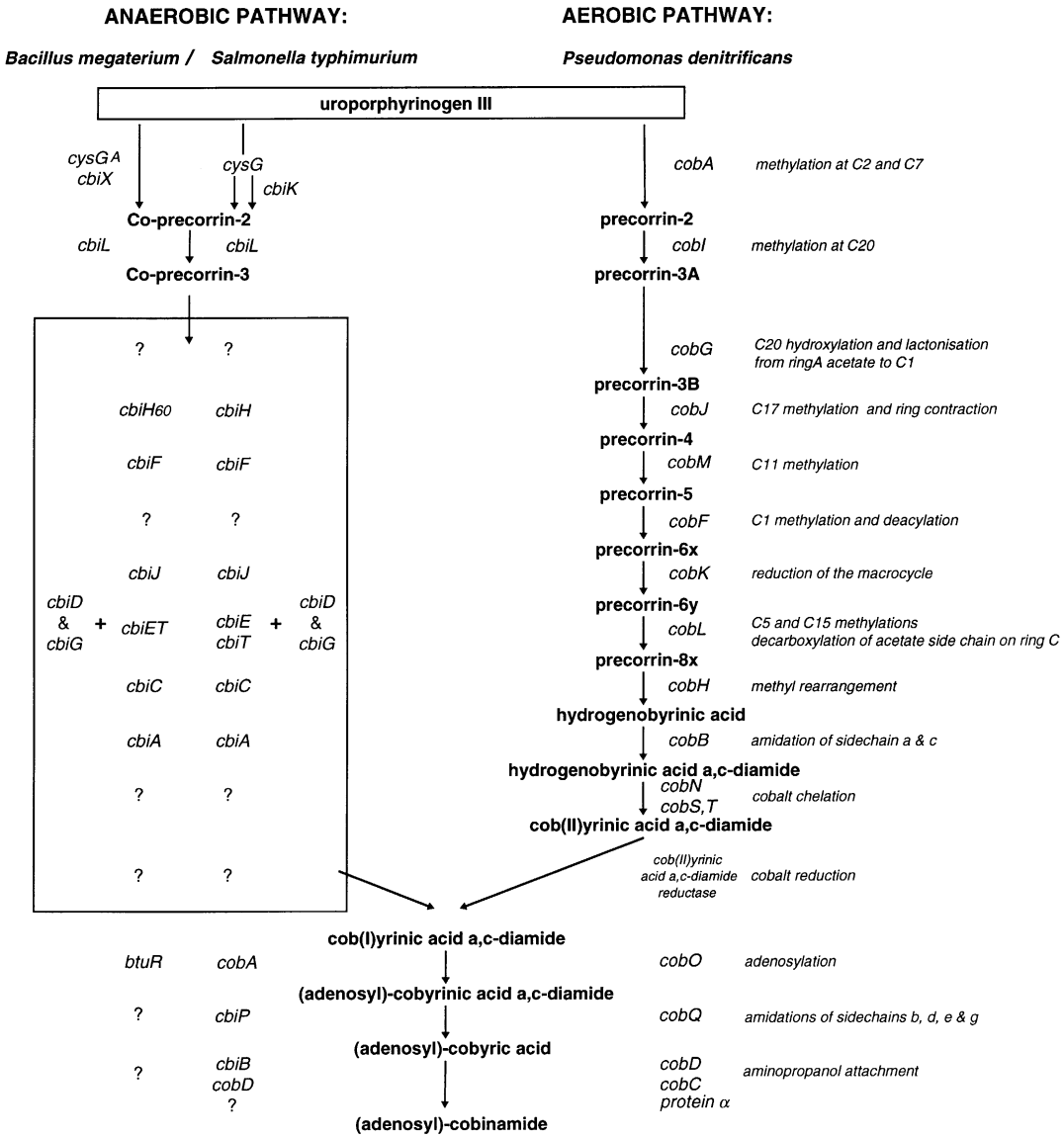
In effect, the molecular genetics of cobalamin biosynthesis have been studied in only these two organisms, *S. typhimurium* and *P. denitrificans* [8–10]. In *P. denitrificans* the genes are dispersed in at least four different regions of the genome, whereas in *S. typhimurium* the genes are located in one major operon. A comparison of the genetic requirements for cobalamin biosynthesis between these two organisms reveals that although a number of the genes display some similarity, a few are quite different (Scheme 1). For instance, *P. denitrificans* does not contain equivalents of CbiD, -G and -K, whereas *S. typhimurium* does not contain equivalents of CobE, -F, -G, -N, -S, -T and -W [8,11]. This reflects the fact that in the synthesis of the corrin component of cobalamin these organisms operate distinct pathways that differ in their requirements for molecular oxygen and the timing of cobalt insertion (Scheme 1). In this respect, the *P. denitrificans* cobalamin pathway employs CobG to generate precorrin-3b with molecular oxygen [7,9,11]. Furthermore cobalt is inserted into hydrogenobyrinic acid *a*,*c*-di-amide by a complex formed between CobN, -S and -T. The cobalamin pathway in *S. typhimurium* chelates cobalt at the level of precorrin-2 in a reaction catalysed by CbiK [12].

Here we describe the main *cob* locus of *Bacillus megaterium* and compare the genes with other recently identified cobalamin biosynthetic sequences from *Methanococcus jannaschi* and *Synechocystis* sp. In comparative terms, *B. megaterium* is regarded as a good producer of cobalamin and has been used in the past for the industrial production of vitamin $B_{12}$. It is likely to act as an excellent model for cobalamin biosynthesis, as a study of the molecular genetics of the cobalamin synthesis genes will provide a detailed understanding of the control and regulation of cobalamin production and give an insight into the biochemistry of the pathway.

Abbreviation used: ORF, open reading frame.

[1] To whom correspondence should be addressed (e-mail e.raux@ucl.ac.uk).

The nucleotide sequence data reported will appear in DDBJ, EMBL and GenBank Nucleotide Sequence Databases under the accession number AJ000758.




**Scheme 1 Outline of the biosynthesis of cobinamide from uroporphyrinogen III**

## MATERIALS AND METHODS

### Chemicals and reagents

Cyanocobalamin, cobinamide and most chemicals were purchased from Sigma. Restriction enzymes and isopropyl β-D-thiogalactoside were purchased from Appligene; Sequenase® T7 DNA polymerase was obtained from USB; calf intestinal alkaline phosphatase and T4 DNA ligase were from Promega; pKK223.3 and Klenow fragment of DNA polymerase I were purchased from Pharmacia; M13mp18 and M13mp19 were obtained from Boehringer; pACYC184 was obtained from New England Biolabs; and [$^{35}$S]dATP, [α-$^{32}$P]dATP and the nick translation kit were obtained from Amersham.

### Bacterial strains, media and growth conditions

*Bacillus megaterium* de Bary DSM 509 and ATCC 10778 were obtained from the appropriate strain collections. All the *E. coli* strains containing plasmids carrying fragments of the *B. megaterium cob* operon were derived from AR3730 (LE392 rendered met$^+$ [11]) and MC1061. The genotypes of all the strains are listed in Table 1. *B. megaterium* was grown aerobically in Luria–Bertani medium to an $A_{600}$ of approx. 1 at 30 °C. Other strains were routinely grown in Luria–Bertani broth or Luria–Bertani agar at 37 °C, with 100 mg/l ampicillin and 30 mg/l chloramphenicol as required. To assess cobalamin production the strains were grown in minimal medium supplemented with 0.1 g/l yeast extract.

### Construction of genomic DNA *B. megaterium* libraries

Genomic DNA from *B. megaterium* DSM 509 was prepared by the method described in [13] and was restricted into 5–10 kb fragments with *Alu*I. The fragments were inserted into the *Sma*I site of pAR8720 (pKK223.3 harbouring the *S. typhimurium cbiA*) to generate the first *B. megaterium* genomic library. This was used to transform *E. coli* strain AR8688, a strain containing plasmid pAR8686 that harbours both the *E. coli cysG* gene and the complete *S. typhimurium cobI* genes, with the exception of

**Table 1 List of strains and plasmids used in this study**

| Strain | Genotype and/or phenotype | Description | Reference or source |
|---|---|---|---|
| ***B. megaterium*** **strains** | | | |
| DSM509 | | | |
| ATCC10778 | | | |
| ***S. typhimurium*** **strain** | | | |
| AR8647 | *S. typhimurium cysG metE* (pBR328) | Phenotypically deficient in the methylase activity of CysG only | This work |
| ***E. coli*** **strain** | | | |
| AR3730 | LE392 *met*$^+$ | | [11] |
| AR8688 | LE392 *met*$^+$ (pAR8686) | Contains *E.c cysG* and all *S.t cbi* genes except *cbiE* and *-A* | [11] |
| MC1061 | F$^-$ araD139 Δ(*ara-Leu*)*7696 galE15 galK16* Δ(*lac*)*X74 rpsL* (Str$^r$) *hsdR2* ($r_K^-$ $m_K^+$)*mcrA mcr B1* | | [11] |

| Plasmid | Inserted genes | Description | |
|---|---|---|---|
| pAR8720 | *S.t cbiA* | | [11] |
| pAR8754 | *S.t cbiA- B.m cbiX-J-C-D-ET* | 6 kb *B. megaterium Alu*I fragment cloned into the *Sma*I site of pAR8720 (from position 3055 to 9063) | This work |
| pAR8766 | *B.m cbiW-H$_{60}$-X-J-C-D-ET-L-F-G-A-cysG$^A$-cbiY-btuR-*> <-ORF1-ORF2 | 16.3 kb *B. megaterium Sau*3AI fragment cloned into *Bam*HI site of pKK223.3 (the *tac* promoter has been deleted) | This work |
| pBR328 | | Vector derived from pBR322; ampicillin, chloramphenicol and tetracycline resistant | Boehringer Mannheim |
| pKK223.3 | | Overexpression vector derived from pBR322 with *tac* promoter; ampicillin resistant | Pharmacia |

*cbiA* and *cbiE*. The resulting colonies obtained from the transformation of AR8688 with the *B. megaterium* genomic library were screened for *cbiE* complementation by the analysis of cobyric acid production in individual colonies on bioassay plates as described below.

### Detection of cobyric acid-synthesizing colonies by bioassay

Bioassay plates were prepared as described previously [11] except that the indicator strain used was AR8647 (*S. typhimurium metE cysG*, AR3612, transformed with pBR328). This strain is able to detect cobyric acid and later intermediates on the cobalamin biosynthetic pathway. Intermediates preceding cobyric acid do not seem to be transported into the reporter strain. The bioassay plates were prepared in minimum medium containing cysteine (50 $\mu$g/ml), ampicillin (50 $\mu$g/ml) and chloramphenicol (35 $\mu$g/ml). Individual clones obtained from the transformation of AR8688 with the *B. megaterium* genomic library were re-streaked on the bioassay plates. Complementation of *cbiE* in AR8688 was observed in one colony by growth of the *S. typhimurium* indicator strain in the agar around the colony after 18–24 h at 37 °C. The plasmid that complemented AR8688 was called pAR8754.

### Second genomic library from *B. megaterium*

A second *B. megaterium* genomic library containing 10–20 kb fragments was generated after a partial *Sau*3AI digestion of *B. megaterium* genomic DNA and ligation of the fragments into *Bam*HI-cut pKK223.3 (thereby removing the *tac* promoter). The library was transformed into *E. coli* MC1061 and was screened for the presence of a cobalamin biosynthetic genes by colony hybridization with the *Hin*dIII fragment from pAR8754 (from *Hin*dIII 8682 to 9061). Before this, the fragment had been labelled with [$\alpha$-$^{32}$P]dATP by nick translation of the DNA in accordance with the instructions of the Amersham nick translation kit. One colony to which the probe annealed contained a plasmid that was called pAR8766.

### Sequencing of the *B. megaterium cob* region

Complete restriction maps of both pAR8754 (harbouring the small *cob* fragment obtained from the first library) and pAR8766 (harbouring the large *cob* fragment obtained from the second library) were generated. The *cob* operon was sequenced after subcloning small fragments from these plasmids into M13 mp18 and mp19 as described previously [11]. For difficult regions with a high degree of secondary structure, dITP or 7-deaza-dGTP were used as required.

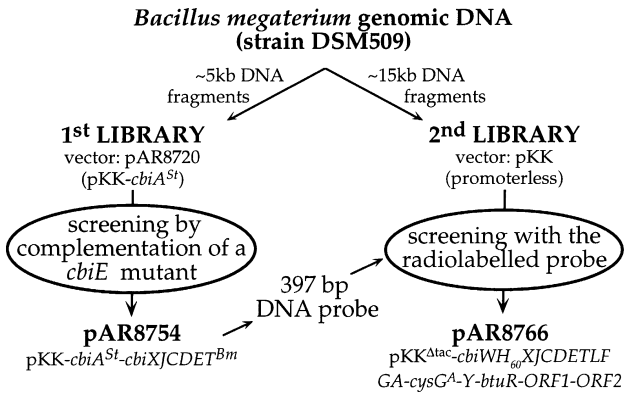
### Computer programs

Sequences, alignments and comparisons were performed with the GCG Wisconsin Package [14]. The provisional detection and assignment of rho-independent terminators was performed with the algorithm described in [15].

## RESULTS AND DISCUSSION

### Isolation of a *B. megaterium* DNA fragment containing *cbi* genes

A *B. megaterium* genomic DNA library containing 5–10 kb DNA fragments was constructed from genomic DNA isolated from strain DSM 509 by ligating restricted DNA fragments into a modified pKK223-3 plasmid that contained the *S. typhimurium cbiA* (pAR8720; Table 1) (Scheme 2). The library was used to transform *E. coli* strain AR8688, which harboured a plasmid containing the *E. coli cysG* and all the *S. typhimurium cbi* genes except for *cbiA* and *-E*. In the absence of *cbiE* the strain (AR8688) is unable to synthesize cobyric acid. Thus it was possible to detect the presence of the *B. megaterium cbiE* by screening for the ability to synthesize cobyric acid after transforming AR8688 with the *B. megaterium* genomic library. As described in the Materials and methods section, this was ac-




**Scheme 2 Outline of the procedure used to isolate the *B. megaterium cobI* operon**

complished by plating the transformed AR8688 on bioassay plates that, on complementation of the *cbiE* deficiency, would promote growth of the *S. typhimurium* indicator strain around clones that produce cobyric acid. The indicator strain relies on a cobalamin-dependent methionine synthase for methionine; growth of the strain is propagated only if cobalamin (or a late intermediate such as cobyric acid) is excreted into the medium. This procedure identified one such clone and the plasmid that led to cobyric acid synthesis was designated pAR8754 (Scheme 2). This plasmid was found to contain a 6 kb DNA insert that was subsequently sequenced fully and, on the basis of sequence similarity, was found to contain five potential *cbi* genes (Figure 1).

To isolate the complete *B. megaterium cob* locus, a second *B. megaterium* genomic library was constructed by ligating larger 10–20 kb *B. megaterium* genomic DNA fragments into *Bam*HI-restricted pKK223-3 (Scheme 2). This procedure also removed the *tac* promoter from the vector (see Table 1). After transformation of this second *B. megaterium* library into *E. coli* MC1061, colonies were screened with a 397 bp radiolabelled DNA probe obtained from the *B. megaterium* region of the previous plasmid, pAR8754. This screening procedure produced one positive clone that harboured a plasmid called pAR8766 containing a DNA fragment of approx. 16 kb. A restriction map of this insert was generated and the DNA sequence of the whole insert was determined.

## Gene organization in the *B. megaterium cob* region

Sequencing of the 16 kb *B. megaterium* fragment harboured within pAR8766 revealed that it contained 15 open reading frames (ORFs), 14 of which were ordered in the same orientation (Figure 1). Of these 14 ORFs, nine had similarity to cobalamin biosynthetic genes found in both *S. typhimurium* and *P. denitrificans*: *cbiH*$_{60}$, *-J*, *-C*, *-ET*, *-L*, *-F*, *-A*, *cysG*$^{A}$ and *btuR* (Figure 1; Table 2). Two ORFs have similarities to *cbi* genes found only in *S. typhimurium* (*cbiD* and *cbiG*), whereas the three remaining ORFs, *cbiW*, *cbiX* and *cbiY*, have no similarity to any known *cob* or *cbi* gene. The ORFs in the *B. megaterium cob* region have been assigned the prefix *cbi*, reflecting the fact that the genes are more similar to those found in *S. typhimurium* than to those in *P. denitrificans*. The exception to this assignment is the last ORF of the operon, which was named *btuR* because of its similarity to the gene found within *E. coli* [16]. Further analysis of these 14 genes strongly suggests that they are co-transcribed and thus constitute a single *cob* operon (Figure 2). There was little or no intercistronic region between the 14 ORFs except for a 1.4 kb region situated upstream of *cbiW*, where no ORF could be detected.

The restriction map of the large 16 kb *cbi* fragment did not correlate with the restriction maps of the cobalamin biosynthetic genes supposedly isolated from the same strain of *B. megaterium* (ATCC 10778) [17,18]. Similarly, the previously reported *cobA* (uroporphyrinogen transmethylase) sequence from *B. megaterium* [19] is only 63% identical at the amino acid level to the equivalent sequence (*cysG*$^{A}$) reported herein. Unless *B. megaterium* has two discrete cobalamin biosynthetic operons, it seems

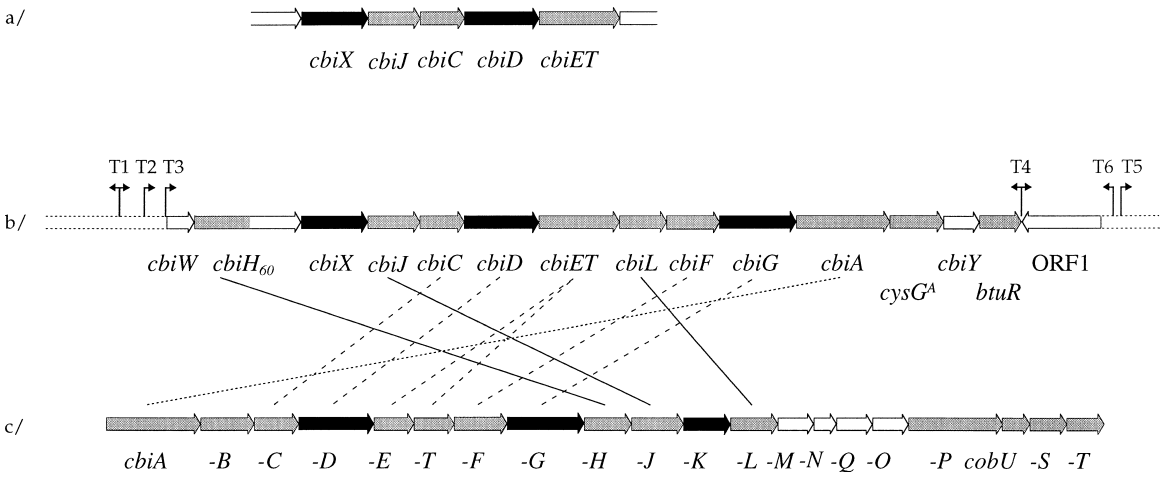



**Figure 1 Schematic representation of the *cobI* gene order found in *B. megaterium* and *S. typhimurium***

(**a**) The order of the *cbi* genes in the small *B. megaterium* genomic DNA fragment cloned into pAR8754. (**b**) The order of the *cobI* genes found in the large *B. megaterium* 16 kb genomic DNA fragment cloned into pAR8766. The presence and direction of the terminations, T1–T6, are also shown. (**c**) The order of the *cbi* genes found in the *S. typhimurium cobI* operon. Similar genes are indicated by the dotted lines. Genes represented by shaded arrows are common to *S. typhimurium*, *B. megaterium* and *P. denitrificans*. Genes represented by black arrows are not found in *P. denitrificans*.

**Table 2 Comparison of *B. megaterium cobI* genes with those from other organisms**

Abbreviations: aa, amino acid; N-ter, N-terminus; C-ter, C-terminus.

| *B. megaterium* | *S. typhimurium* | | *P. denitrificans* | | *M. jannaschii* | | *Synechocystis* sp. | |
|---|---|---|---|---|---|---|---|---|
| Gene | Aa identity (%) | Gene | Aa identity (%) | Gene | Aa identity (%) | ORF | Aa identity (%) | ORF |
| *cbiW* | | | | | | | 30.3 | sll1584 |
| $cbiH_{60}$ (N-ter) | 51.7 | *cbiH* | 37.8 | *cobJ* | 43.1 | MJ0813 | 53.9 | slr0969 (C-ter) |
| (C-ter) | | | | | 22.4 | MJ0551 | 25.5 | slr0898 |
| *cbiX* | | | | | 25.3 | MJ0970 | 36.2 | sll0037 |
| *cbiJ* | 33.2 | *cbiJ* | 26.5 | *cobK* | 32.1 | MJ0552 | 29.1 | slr0252 |
| *cbiC* | 39.1 | *cbiC* | 34.4 | *cobH* | 43.9 | MJ0930 | 37.2 | sll0916 |
| *cbiD* | 34.4 | *cbiD* | | | 37.5 | MJ0022 | 24.9 | slr1538 |
| *cbiET* | 33.5 | *cbiE* | 32.4 | *cobL* | 32.5 | MJ1522 (*E*) | 41.7 | sll0099 (*ET*) |
| | 32.6 | *cbiT* | | | 35.4 | MJ0391 (*T*) | 32.4 | slr1368 (*T*) |
| *cbiL* | 29.8 | *cbiL* | 27.0 | *cobI* | 34.3 | MJ0771 | 35.8 | slr1879 |
| *cbiF* | 43.8 | *cbiF* | 38.2 | *cobM* | 52.5 | MJ1578 | 43.3 | slr0239 |
| *cbiG* | 30.8 | *cbiG* | 28.1 | *cobE* (C-ter of *cbiG*) | 24.3 | MJ1144 | 30.4 | slr0969 (N-ter) |
| *cbiA* | 37.3 | *cbiA* | 36.4 | *cobB* | 40.0 | MJ1421 | 38.1 | sll1501 |
| $cysG^A$ | 47.1 | *cysG* (C-ter) | 45.0 | *cobA* | 55.2 | MJ0965 | 50.0 | sll0378 |
| *cbiY* | | | | | | | | |
| *btuR* | 36.1 | *cobA* | 37.5 | *cobO* | | | 37.7 | slr0260 |

| Gene | Overlap of stop/start codons | position of First codon | position of Stop codon | No. of amino acids |
|---|---|---|---|---|
| *cbiW* | AAAGGAGCGA **ATG** AGT **GTG** | 1764 | 2145 | 127 |
| $cbiH_{60}$ | TAAAGGAGGAGCTAGC **ATG** | 2161 | 3781 | 540 |
| *cbiX* | GGAGATTAAGAA **ATG** GGAGGACATTAT **ATG** | 3787 | 4705 | 306 |
| *cbiJ* | GGCGAGCTAAA **ATG** A | 4704 | 5469 | 255 |
| *cbiC* | AGGAGTTGATTGAG **ATG** | 5477 | 6152 | 225 |
| *cbiD* | AGGTGTAAAACGT **ATG** AAGGAAGTGCCAAAAGAACCTAA | 6129 | 7230 | 367 |
| *cbiET* | AGGCGGTGCAAC **ATG** GCAATTAA | 7222 | 8416 | 398 |
| *cbiL* | AAAAGAGGAGA **ATG** AAC **ATG** | 8415 | 9120 | 235 |
| *cbiF* | GGTGGTGCGAAA **ATG** A | 9119 | 9893 | 258 |
| *cbiG* | AAGGGAGTGAAATCGGA **ATG** A | 9892 | 11014 | 374 |
| *cbiA* | TAAAGGAGAACAAC **ATG** | 11030 | 12410 | 460 |
| $cysG^A$ | AAGAAGAGATGCAGT **GTG** A | 12409 | 13138 | 243 |
| *cbiY* | TAA...37bp...AAGGAGGAGACAAC **ATG** | 13192 | 13759 | 189 |
| *btuR* | AATGAAAGAGC **CTG** CTAGAACACCGTTTACAGAAAAAAC AACATGGCTGTAA .....TAA | 13721 | 14330 | 203 |
| ORF1 | GTTAACATATAAG**ATG**...TAA | 15483 | 14361 | 374 |

**Figure 2 Display of the *B. megaterium cobI* genes codon start/stop signals**

The start codon for the gene is highlighted in bold; the stop signal for the previous gene is underlined. The positions of the first and last codons are also given together with the predicted number of amino acid residues in the protein. For *cbiW*, *cbiX* and *cbiL* there are two possible start codons, each with their own potential ribosome-binding site. In these cases the position of the first codon is taken from the first of the two potential start codons.

likely that one of the strains was not what it was reported to be. To confirm that the *cob* operon described in this report was not derived from some contaminating DNA, fresh genomic DNA preparations from *B. megaterium* strains ATCC 10778 and DSM 509 were used as template for the amplification of $cysG^A$ (*cobA*) by PCR. The amplified product was cloned, sequenced and found to be identical to the *B. megaterium* $cysG^A$ sequence reported herein. Thus the differences in the restriction map of the main *cob* operons reported in [17] and the *B. megaterium cobA* gene sequence reported in [19] must be due either to the presence

of a second *cob*/sirohaem operon or, to the isolation of a *cob* operon from a related *Bacillus* species.

### *B. megaterium cbi* genes

On the basis of sequence similarity a number of the *B. megaterium cbi* genes can be assigned functionality. For instance, the operon contains five transmethylases ($CysG^A$, CbiE, -L, -F and $-H_{60}$) that also display some similarity between themselves. This similarity is greatest around the putative *S*-adenosyl-L-methionine-binding motif Gly-Xaa-Gly-Xaa-Gly [20] and reflects the probable evolutionary origin of the methyltransferases as discussed by [8]. Very recent work on the structure determination of *B. megaterium* CbiF reveals that these cobalamin biosynthetic transmethylases represent a new class of transmethylases [21].

A further point concerning possible *S*-adenosyl-L-methionine-binding site motifs is that another glycine-rich sequence is found within CbiD (Gly-Gly-Xaa-Gly-Xaa-Gly at positions 100–110). During cobalamin biosynthesis, eight methylations of the tetrapyrrole framework occur, which are mediated by six enzymes [7,10]. However, although all six enzymes have been identified in *P. denitrificans* (*cobA*, *-I*, *-J*, *-F*, *-M* and *-L*) [10], only five of the enzymes have been identified in *S. typhimurium* (*cysG*, *cbiL*, *-H*, *-F* and *-E*) [8]; no homologue to *P. denitrificans cobF* has been found. Although *cbiD* is essential for cobalamin biosynthesis in both *S. typhimurium* and *B. megaterium*, no functional role has been ascribed to the protein. It is therefore possible that CbiD might have the same role as CobF in undertaking the C-1 methylation and deacylation reactions required during the ring contraction process.

The genes of the *B. megaterium cobI* operon do, however, differ from their *S. typhimurium* counterparts in a number of ways. First, *B. megaterium* contains a gene that represents a fusion between *cbiE* and *cbiT*, similar to the situation found in *P. denitrificans* (*cobL*). Secondly, the *B. megaterium* $cbiH_{60}$ was considerably longer than that found in *S. typhimurium* (540 amino acid residues compared with 241). The similarity was contained within the first 250 residues of the *B. megaterium* gene, whereas the C-terminal 290-residue sequence was not similar to any known cobalamin biosynthetic gene, although some similarity has been detected to a C-terminal region of a *Synechococcus* sp. ferredoxin:nitrite reductase (28.8% identity) [22], including four conserved cysteine residues in a region proposed as 4Fe-4S/sirohaem domain found in nitrite and sulphite reductases [23]. The role of the latter half of $CbiH_{60}$ in cobalamin biosynthesis is therefore unknown but might relate to the regulation of cobalamin biosynthesis in *B. megaterium*.

Finally, *B. megaterium* contains three genes that are found in neither the *cob* genes of *P. denitrificans* nor the *cbi* genes of *S. typhimurium*, CbiX, -Y and -Y. CbiX is striking in that it contains a histidine-rich region at its C-terminus, which suggests that it might be involved in metal chelation [24]. A similar polyhistidine motif has also been observed in HypB, encoded within part of the operon for hydrogenase synthesis, and is thought to be involved in binding nickel [25,26]. In contrast, CbiW has some similarity to a *Clostridium pasteurianum* ferredoxin [27] and, to a smaller extent, to a *Synechocystis* sp. ORF [28,29]. Interestingly, these proteins contain four conserved cysteine residues. As it is known that during corrin biosynthesis the modified tetrapyrrole undergoes an oxidation–reduction process during ring contraction, which requires molecular oxygen in *P. denitrificans*, it is possible that CbiW could act as oxidoreductant during the ring contraction process under anaerobic conditions. Finally, CbiY presents some similarity to an *E. coli* ORF with an unknown function [30], and to a *Vibrio fischeri* NAD(P):flavin oxidoreductase [31].

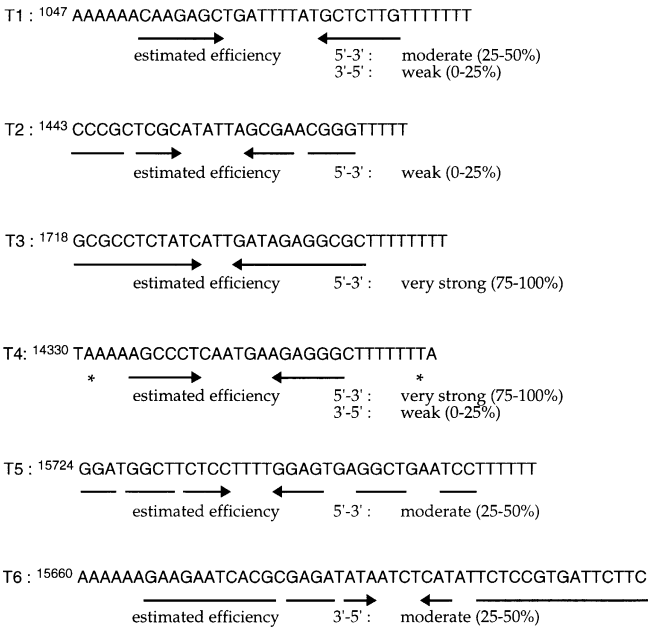



**Figure 3 Primary structure of the terminators, T1–T6, identified by the algorithm described in [15], which are located on each side of the *B. megaterium cobI* operon**

The gene order within the *B. megaterium cob* operon shows some similarity to that within the *S. typhimurium cob* operon, as outlined in Figure 1. Previous studies of gene order among bacteria has shown that equivalent genes are not conserved when the bacteria are distantly related [32]. In the *cob* operons a greater conservation of gene order is observed between the Gram-negative *S. typhimurium* and the Gram-positive *B. megaterium* than between the two Gram-negative bacteria *S. typhimurium* and *P. denitrificans*.

### The *B. megaterium cbi* genes constitute an operon

The *B. megaterium cob* locus identified in this study is putatively constituted of 14 genes, of which at least 11 are expressed when the locus is inserted into a promoterless plasmid. This is demonstrated by the ability of the *E. coli* strain ER241 to synthesize cobyric acid (550–600 pmol per attenuance unit in the the presence or absence of isopropyl $\beta$-D-thiogalactoside) even though the plasmid containing the main *cob* operon (pAR8766) does not contain the *tac* promoter (the following paper [33]). Thus the genes are probably transcribed from promoter(s) situated in the 1.7 kb DNA fragment upstream of *cbiW*. A search of the 16 kb *B. megaterium cob* DNA fragment with the use of the algorithm described in [15] for rho-independent terminators revealed the presence of six separate termination signals (T1–T6); their location and sequences are given in Figures 1 and 3. Termination signals T1–T3 are located upstream of the first gene of the operon, whereas signals T4–T6 are located downstream of *btu*R. Thus the 14 contiguous genes are sandwiched between the strong termination signals T3 and T4. No other signal was found between T3 and T4. The positions of these termination signals suggest that the 14 genes from *cbiW* to *btuR* constitute a single transcription unit. The position of the T3

terminator immediately upstream of *cbiW* indicates that T3 is very likely to be part of a transcription regulation region. When repressed by high concentrations of endogenous cobalamin, as observed with the *S. typhimurium* CobI–III region [34], the T3 structure could hypothetically form on the transcribed RNA and block the RNA polymerase elongation, as has been described in a number of attenuation systems [35]. When derepressed, perhaps when cobalamin concentrations are very low, interactions of the cobalamin molecule with the RNA regulatory region would not permit the T3 structure to form, and would allow active transcription of the *cob* operon.

### Comparison of the *B. megaterium cbi* genes with the cobalamin biosynthetic genes identified from genome sequencing projects

A number of ORFs from bacterial genome sequencing projects have been identified as cobalamin biosynthetic genes through sequence similarity to either the *P. denitrificans cob* or the *S. typhimurium cbi* genes. In this section the *B. megaterium* Cbi protein sequences are compared with the putative cobalamin biosynthetic protein sequences that have been identified from the genome sequencing projects of *Methanococcus jannaschii* [36] and *Synechocystis* sp. [28,29]. However, it should be kept in mind that proteins with no sequence similarity can replace each other in the same pathway, as has been shown recently for the *S. typhimurium* $CysG^{B}$ and CbiK [12].

### *M. jannaschii* cobalamin pathway

Unlike *S. typhimurium* and *B. megaterium*, the cobalamin biosynthetic genes of *M. jannaschii* are not organized into operons but are dispersed throughout the genome. However, on the basis of sequence similarity 17 genes have been identified that might be involved in the transformation of uroporphyrinogen III into cobalamin (see Table 2).

Surprisingly, two different *cbiM* genes are identified in the *M. jannaschii* database (MJ1091 and MJ1569); the gene product referenced MJ1091 (233 residues) has 51% identity in a 206-residue overlap with *S. typhimurium cbiM*, whereas the MJ1569 gene product has only 30.6% identity in a 170-residue overlap. MJ1091 *cbiM* is found adjacent to *cbiN*, *-Q* and *-O* (MJ1090-1089 and MJ1088), in the same gene order as that observed in *S. typhimurium*. This small gene bunch is the only significant clustering of any of the cobalamin bioynthetic genes within *M. jannaschii*. Furthermore no equivalent to the *E. coli btuR*, which catalyses the adenosylation of the corrin molecule [16], is found in the *M. jannaschii* genome. This suggests that the organism does not require adenosylcobalamin. The methylcobalamin form is required as a coenzyme in the process of methanogenesis.

### *Synechocystis* sp. cobalamin biosynthetic pathway

As with *M. jannaschii*, the cobalamin biosynthetic genes in *Synechocystis* sp. are not clustered into operons. Nevertheless, on the basis of sequence similarity 21 ORFs have been identified as cobalamin biosynthetic genes (Table 2). The *Synechocystis* sp. genome contains two copies of *cbiT* (precorrin-8W decarboxylase); ORF sll0099 is designated as *cbiE* but actually encodes a bifunctional enzyme CbiET corresponding to *S. typhimurium* CbiE and CbiT. This is equivalent to the situation with both *P. denitrificans* CobL and *B. megaterium* CbiET. The second copy of *cbiT* is found within ORF slr1368, but why *Synechocystis* requires two versions of CbiT is unclear.

The gene encoded by ORF slr0969 has been designated *cbiH* but in fact represents a fusion between *cbiH* and *cbiG* and should therefore be termed *cbiGH*. The encoded protein consists of 627 residues; the C-terminal portion of the protein (residues 350–627) displays similarity to CbiH, and the N-terminal portion to CbiG. As other multifunctional proteins involved in cobalamin biosynthesis catalyse adjacent steps in the pathway, including CysG [37], CobL (CbiET) [10], CobIJ [38] and CobA-HemD [39], it is therefore possible that CbiG catalyses a reaction step adjacent to CbiH. In the anaerobic pathway such a step could be the formation of a gamma lactone, which is thought to help to mediate the anaerobic ring contraction process [40].

### *M. jannaschii* and *Synechocystis* sp. lack the genetic hallmarks of the aerobic cobalamin pathway

Of particular significance in the genomes of *M. jannaschii* and *Synechocystis* sp. is the absence of homologues to *P. denitrificans cobG* and *cobF*, whose products encode the molecular oxygen requiring C-20 hydroxylase and the C-1 methylase respectively. Nevertheless both genomes are reported to contain putative CobN proteins (MJ0908, slr1211 and sll1870), identified on the basis of their similarity to the *P. denitrificans* CobN, a component of the CobN, -S, -T complex [10] which inserts cobalt at a late stage in the aerobic cobalamin biosynthetic pathway. However, $CobN^{Pd}$ is also similar to a number of other proteins, including an *M. jannaschii* Mg-chelatase, and a *Synechocystis* sp. Mg-chelatase. Consequently, although it is likely that the CobN proteins (MJ0908, slr1211 and sll1870) are chelatases, it is not clear whether they are cobaltochelatases. For instance, MJ0908 could act as a nickel chelatase in coenzyme $F_{430}$ synthesis. Furthermore, as no CobS or -T homologues are found within either genomes, the other components of the *P. denitrificans* cobaltochelatase complex, it is unlikely that the proteins could act as cobaltochelatases as described in *P. denitrificans* [10].

Taken together, the presence in the *M. jannaschii* and *Synechocystis* sp. genomes of genes coding for CbiD, -G and -X, and the absence of CobG and CobF, strongly suggest that these organisms operate an anaerobic cobalamin pathway. This pathway is characterized by the early insertion of cobalt, at the level of precorrin-2, and the absence of a requirement for molecular oxygen. The *B. megaterium* operon also contains *cbiX*, which encodes for a protein that is necesssary for cobyric acid synthesis when the *B. megaterium cob* operon is expressed in *E. coli*; similarly the *B. megaterium cob* operon does not contain an equivalent to *cbiK*. This clearly shows that the *B. megaterium* and *S. typhimurium* pathways, although very similar, are not identical. The sequencing of the cyanobacterium *Synechocystis* sp. genome revealed that this organism possesses cobalamin biosynthetic genes similar to *cbiD*, *-G* and *X*, relating this $B_{12}$ pathway to the *B. megaterium* $B_{12}$ pathway. By contrast, the cobalamin pathway in the archaebacterium *M. jannaschii*, which exhibits a *cbiD* gene and possibly *cbiG* but does not contain either *cbiX* or *cbiK*, might represent yet another type of anaerobic pathway.

Financial support from the BBSRC is gratefully acknowledged.

## REFERENCES

1 Beck, W. S. (1982) in $B_{12}$, vol. 2 (Dolphin, D., ed.), pp. 1–30, Wiley, New York
2 Schneider, Z. and Stroinski, A. (1987) in Comprehensive $B_{12}$ (Schneider, Z. and Stroinski, A., eds.), pp. 259–266, Walter de Gruyer, Berlin
3 Stroinski, A. (1987) in Comprehensive $B_{12}$ (Schneider, Z. and Stroinski, A., eds.), pp. 226–259, Walter de Gruyer, Berlin

4 Warren, M. J. and Scott, A. I. (1990) Trends Biochem. Sci. **15**, 486–491
5 Benner, S. A., Ellington, A. D. and Traver, A. (1989) Proc. Natl. Acad. Sci. U.S.A. **86**, 7054–7058
6 Eschenmoser, A. (1988) Angew. Chem. Int. Ed. Engl. **27**, 5–39
7 Scott, A. I. (1994) Tetrahedron **50**, 13315–13333
8 Roth, J. R., Lawrence, J. G., Rubenfield, M., Kieffer-Higgins, S. and Church, G. M. (1993) J. Bacteriol. **175**, 3303–3316
9 Debussche, L., Thibaut, D., Cameron, B., Crouzet, J. and Blanche, F. (1993) J. Bacteriol. **175**, 7430–7440
10 Blanche, F., Cameron, B., Crouzet, J., Debussche, L., Thibaut, D., Vuilhorgne, M., Leeper, F. J. and Battersby, A. R. (1995) Angew. Chem. Int. Ed. Engl. **34**, 383–411
11 Raux, E., Lanois, A., Levillayer, F., Warren, M. J., Brody, E., Rambach, A. and Thermes, C. (1996) J. Bacteriol. **178**, 753–767
12 Raux, E., Thermes, C., Heathcote, P., Rambach, A. and Warren, M. J. (1997) J. Bacteriol. **179**, 3202–3212
13 Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual, 2nd edn., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
14 Program Manual for the Wisconsin Package, Version 8, September 1994, Genetics Computer Group, 575 Science Drive, Madison, WI
15 D'Aubenton Carafa, Y., Brody, E. and Thermes, C. (1990) J. Mol. Biol. **216**, 835–858
16 Lundrigan, M. D. and Kadner, R. J. (1989) J. Bacteriol. **171**, 154–161
17 Brey, R. N., Banner, C. D. B. and Wolf, J. B. (1986) J. Bacteriol. **167**, 623–630
18 Wolf, J. B. and Brey, R. N. (1986) J. Bacteriol. **166**, 51–58
19 Robin, C., Blanche, F., Cauchois, L., Cameron, B., Couder, M. and Crouzet, J. (1991) J. Bacteriol. **173**, 4893–4896
20 Warren, M. J., Bolt, E. and Woodcock, S. C. (1994) in The Biosynthesis of the Tetrapyrrole Pigments (Ciba Found. Symp. **180**) (Chadwick, D. J. and Ackrill, K., eds.), pp. 26–40, Wiley, Chichester
21 Schubert, H., Wilson, K. S., Raux, E., Woodcock, S. C. and Warren, M. J. (1998) Nature Struct. Biol. **5**, 585–593
22 Luque, I., Flores, E. and Herrero, A. (1993) Plant Mol. Biol. **21**, 1201–1205
23 Crane, B. R., Siegel, L. M. and Getzoff, E. D. (1995) Science **270**, 59–67
24 Beck, R., Raux, E., Thermes, C., Rambach, A. and Warren, M. J. (1997) Biochem. Soc. Trans. **25**, 77S
25 Rey, L., Imperial, J., Palacios, J. M. and Ruizargueso, T. (1994) J. Bacteriol. **176**, 6066–6073
26 Fu, C. L., Olson, J. W. and Maier, R. J. (1995) Proc. Natl. Acad. Sci. U.S.A. **92**, 2333–2337
27 Meyer, J. (1993) Biochim. Biophys. Acta **1174**, 108–110
28 Kaneko, T., Tanaka, A., Sato, S., Kotani, H., Sazuka, T., Miyajima, N., Sugiura, M. and Tabata, S. (1995) DNA Res. **2**, 153–166
29 Kaneko, T., Sato, S., Kotani, H., Tanaka, A., Asamizu, E., Nakamura, Y., Miyajima, N., Hirosawa, M., Sugiura, M., Sasamoto, S. et al. (1996) DNA Res. **3**, 109–136
30 Sawers, G., Heider, J., Zehelein, E. and Bock, A. (1991) J. Bacteriol. **173**, 4983–4993
31 Zenno, S., Saigo, K., Kanoh, H. and Inouye, S. (1994) J. Bacteriol. **176**, 3536–3543
32 Mushegian, A. and Koonin, E. V. (1996) Trends Genet. **12**, 289–290
33 Raux, E., Lanois, A., Rambach, A., Warren, M. J. and Thermes, C. (1998) Biochem. J. **335**, 167–173
34 Richter-Dahlfors, A. A. and Andersson, D. I. (1992) J. Mol. Biol. **6**, 743–749
35 Landick, R. and Yanofsky, C. (1987) in *E. coli* and *S. typhimurium* Cellular and Molecular Biology (Neidhardt, F. C., ed.), pp. 1276–1301, American Society for Microbiology, Washington, DC
36 Bult, C. J., White, O., Olsen, G. J., Zhou, L., Fleischmann, R. D., Sutton, G. G., Blake, J. A., FitzGerald, L. M., Clayton, R. A., Gocayne, J. D. et al. (1996) Science **273**, 1058–1073
37 Warren, M. J., Bolt, E. L., Roessner, C. A., Scott, A. I., Spencer, J. B. and Woodcock, S. C. (1994) Biochem. J. **302**, 837–844
38 Philipp, W. J., Poulet, S., Eiglmeier, K., Pascopella, L., Balasubramanian, V., Heym, B., Bergh, S., Bloom, B. R., Jacobs, W. R. and Cole, S. T. (1990) Proc. Natl. Acad. Sci. U.S.A. **93**, 3132–3137
39 Fujino, E., Fujino, T., Karita., S., Sakka, K. and Ohmiya, K. (1995) J. Bacteriol. **177**, 5169–5175
40 Scott, A. I., Stolowich, N. J., Wang, J. J., Gawatz, O., Fridrich, E. and Muller, G. (1996) Proc. Natl. Acad. Sci. U.S.A. **93**, 14316–14319

Received 13 March 1998; accepted 31 July 1998